

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 13:49:28 ; Search time 73 seconds  
(without alignments)  
1038.625 Million cell updates/sec

Title: US-09-805-337A-2

Perfect score: 3183

Sequence: 1 MLLLFVILLISWTVGGEG.....SSPPFRAICQEGKFEYPICE 569

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3183	100.0	569	AAE23195	Human factor H-rel
2	1691	53.1	327	ABH80571	Human sbg614126com
3	1599	50.2	364	ABH80570	Human sbg614126com
4	1492	46.9	267	ABG66700	Human novel polype
5	1303	40.9	678	ABP41931	Human ovarian anti
6	1051.5	33.0	342	AAH43738	Human cancer assoc
7	1001.5	31.5	578	AAH09065	Human complement f
8	756.5	23.8	660	AAH82921	B subunit of human
9	708.5	22.3	582	AAU02929	Angiotensin conver
10	562	17.7	290	AAW39156	Human partial comp

11	466	14.6	229	18	AAW39158	Clone PRBS3FH2910
12	432.5	13.6	196	18	AAW39159	Clone PRBS3FH2910
13	408	12.8	1087	12	AAW11139	B lymphocyte membr
14	395	12.4	979	22	AAW66398	Receptor 222. Uni
15	395	12.4	1066	22	ABP37987	Human GS96663 prot
16	395	12.4	1139	22	ABU87355	Novel central nerv
17	395	12.4	1139	22	ABH10326	Human CDNA SEQ ID
18	395	12.4	1139	22	AAU18126	Novel human uterin
19	395	12.4	1139	22	AAU16963	Human novel secret
20	395	12.4	1139	22	AAU19902	Novel human calciu
21	395	12.4	1251	23	AAU74829	Human REPTR 12 pro
22	394	12.4	3571	23	AAE20146	Human C3b/C4b comp
23	389.5	12.2	330	22	AAU02952	Angiotensin conver
24	389	12.2	2050	23	AAW68264	Human POLY11 prote
25	385	12.1	1537	12	AAW11982	Partial human comp
26	385	12.1	1929	22	ABG00103	Novel human diagno
27	385	12.1	1930	19	AAW45899	Human complement r
28	385	12.1	2039	12	AAW11810	Human complement t
29	385	12.1	2039	20	AAW55751	Human C3b/C4b rece
30	385	12.1	2039	22	ABG00287	Novel human diagno
31	385	12.1	2039	22	AAW50797	Human C3B/C4B rece
32	385	12.1	2044	22	AAW39224	Human polypeptide
33	385	12.1	2317	10	AAW92219	CR1 protein. Homo
34	384.5	12.1	3594	23	AAE20147	Mouse C3b/C4b comp
35	383.5	12.0	841	23	ABP51418	Human MDR1 SEQ ID
36	382	12.0	1497	22	AAW93954	Human polypeptide,
37	380.5	12.0	778	19	AAW73147	Amino acid sequenc
38	380.5	12.0	2044	22	ABH11782	Human CR1 protein
39	380.5	12.0	2044	22	AAW41010	Human polypeptide
40	379.5	11.9	645	21	AAW33125	Macaca mulatta rha
41	376	11.8	2039	14	AAW36743	CR1. Homo sapiens
42	370	11.6	3069	23	AAE20787	Human C3b/C4b comp
43	370	11.6	3069	23	AAE20900	Human C3b/C4b comp
44	370	11.6	3100	23	AAE20789	Human C3b/C4b comp
45	370	11.6	3100	23	AAE20901	Human C3b/C4b comp

## ALIGNMENTS

RESULT 1	
AAE23195	
ID	AAE23195 standard; Protein; 569 AA.
XX	
AC	AAE23195;
XX	
DT	21-AUG-2002 (first entry)
XX	
DE	Human factor H-related protein (FHR-5).
XX	
KW	Human; factor H-related protein; FHR-5; detection; C5b-9 complement;
KW	immunohistochemical; diagnostic; immunological; biopsy; prevention.
XX	
OS	Homo sapiens.
XX	
PN	WO200168695-A2.
XX	
PD	20-SEP-2001.
XX	
PF	13-MAR-2001; 2001WO-US07868.
XX	
PR	13-MAR-2000; 2000US-0188870.
XX	
PA	(BAXT ) BAXTER INT INC.
XX	
PI	Murphy BF;
XX	
DR	WPI; 2001-582437/65.
DR	N-PSDB; AAD37114.
XX	
PT	New factor-H related protein 5 that binds to complement component C3b,
PT	useful for raising antibodies suitable for detection of C5b-9
PT	complement complexes,



QY 241 FIINGPKKIOCVGEGWTLPTCEVQKTCGYIPELEYGVQSPVPPYQHGVSVEVNCNE 300  
Db 144 -----MKTCTGYIPELEYGVQSPVPPYQHGVSVEVNCNE 178  
QY 301 YAMIGNNMTICINGIWTLPVCMVATHQLKCKIAGVNIKTLLKLSGKEFNHNSRIRYRCS 360  
Db 179 YAMIGNNMTICINGIWTLPVCMVATHQLKCKIAGVNIKTLLKLSGKEFNHNSRIRYRCS 360  
QY 361 DIFRYHVSVCINGKWNPEVDCTEKREQFCPPPPQIPNAQNMNTTVYQDGEKVAVLCKEN 420  
Db 200 ----- 199  
QY 421 YLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTTSFPLSVYPPGSTVYRQCS 480  
Db 200 -----CVSTAYCGPPPSINNGDTTSFPLSVYPPGSTVYRQCS 238  
QY 481 FYKLGQSVTVTCRNKQWSEPPRCIDPCVSVSENNMKNKIQLKWRNDGKLYAKTGDAVEFQ 540  
Db 239 FYKLGQSVTVTCRNKQWSEPPRCIDPCVSVSENNMKNKIQLKWRNDGKLYAKTGDAVEFQ 298  
QY 541 CKFPHKAMISSPPFRAICQEGKFEYPICE 569  
Db 299 CKFPHKAMISSPPFRAICQEGKFEYPICE 327

RESULT 3

ABB80570  
ID ABB80570 standard; Protein; 364 AA.  
XX AC ABB80570;  
XX DT 08-OCT-2002 (first entry)  
XX DE Human sbg614126complifh protein #1.  
XX KW Human; secreted protein; immunosuppressive; cytostatic; nootropic;  
KW neuroprotective; antitumour; vulnary; antimicrobial; ophthalmological;  
KW antiparkinsonian; antirheumatic; antiatherosclerotic; dermatological;  
KW hypotensive; cerebroprotective; virucide; antiinflammatory; diabetes;  
KW malignant tumour; hypertension; hypotension; obesity; bulimia; anorexia;  
KW asthma; manic depression; dementia; delirium; mental retardation;  
KW Huntington's disease; Tourette's syndrome; schizophrenia;  
KW mental disorder; sexual development disorder; blood cascade dysfunction;  
KW stroke; growth disorder.  
XX OS Homo sapiens.  
XX PN WO200222802-A1.  
XX PD 21-MAR-2002.  
XX PF 13-SEP-2001; 2001WO-US28462.  
XX PR 13-SEP-2000; 2000US-232455P.  
XX PR 13-SEP-2000; 2000US-232453P.  
XX PR 02-OCT-2000; 2000US-237293P.  
XX PR 07-NOV-2000; 2000US-246269P.  
XX PR 20-NOV-2000; 2000US-252049P.  
XX PA (SMK ) SMITHKLINE BEECHAM CORP.  
XX PA (SMK ) SMITHKLINE BEECHAM PLC.  
XX PA (GLAX ) GLAXO GROUP LTD.  
XX PI Agarwal P, Gogswell JP, Lai Y, Martensen SA, Smith RF, Strum JC;  
PI Xie Q;  
XX DR WPI; 2002-393963/42.  
XX DR N-PSDB; ABL57706.  
XX PT Novel isolated secreted polypeptides and polynucleotides encoding them  
XX PT useful for treating cancer, Alzheimer's disease, tumor metastasis,  
XX PT autosomal recessive atypical hemolytic uremic syndrome, wound healing

disorder -  
PS 1: Page191-192; 246pp; English.  
XX The invention relates to a novel isolated polypeptide (ABB80569-ABB80612)  
CC (secreted polypeptide) which is encoded by any one of 44 polynucleotide  
CC sequences (ABL57705-ABL57748) given in the specification. The  
CC polypeptides have immunosuppressive, cytostatic, nootropic,  
CC neuroprotective, antitumour, vulnary, antimicrobial, ophthalmological,  
CC antiparkinsonian, antirheumatic, antiatherosclerotic, dermatological,  
CC hypotensive, cerebroprotective, virucide, and antiinflammatory activity.  
CC The polynucleotide and polypeptide are useful for treating diabetes,  
CC malignant tumours, hyper- and hypotension, obesity, bulimia, anorexia,  
CC asthma, manic depression, dementia, delirium, mental retardation,  
CC Huntington's disease, Tourette's syndrome, schizophrenia, growth, mental  
CC or sexual development disorders, and dysfunctions of the blood cascade  
XX system including those leading to stroke.  
SQ Sequence 364 AA;  
Query Match 50.2%; Score 1599; DB 23; Length 364;  
Best Local Similarity 55.8%; Pred. No. 3.5e-120;  
Matches 307; Conservative 1; Mismatches 0; Indels 242; Gaps 2;  
QY 20 GTLCDPFIHHGFLYDEEDYNPFQVPTGEVYFYSCYFNFSVPSKSFTRITCTEGWSP 79  
Db 57 GTLCDPFIHHGFLYDEEDYNPFQVPTGEVYFYSCYFNFSVPSKSFTRITCTEGWSP 116  
QY 80 TPKCLRMCSFPFVKNGHSESSGLIHLEGDVTQIICNTGYSYLNNEKNISCVRGWSTPPI 139  
Db 117 TPKCLRMCSFPFVKNGHSESSGLIHLEGDVTQIICNTGYSYLNNEKNISCVRGWSTPPI 176  
QY 140 CSFTKGECHVPFILEANVDAQPKESYKVGVDVLKFSCKNLIRVSDSVQCYQFGWSPNP 199  
Db 177 CSFT----- 180  
QY 200 TKGQVRSCGPPPPQLSNGEYKEIRKEEYGHNEWEYDCNPNFIINGPKKIQCVGEGWTL 259  
Db 181 ----- 180  
QY 260 PTCVQVQVTCGYIPELEYGVQSPVPPYQHGVSVEVNCNEYAMIGNNMTICINGIWT 319  
Db 181 -----MKTCTGYIPELEYGVQSPVPPYQHGVSVEVNCNEYAMIGNNMTICINGIWT 234  
QY 320 PWCVATHQLKCKIAGVNIKTLLKLSGKEFNHNSRIRYRCSDFRYRHSVCINGKWNPEV 379  
Db 235 PM----- 236  
QY 380 DCTEKREQFCPPPPQIPNAQNMNTTVYQDGEKVAVLCKENYLLPEAKEIVCKDGRWQSL 439  
Db 237 ----- 236  
QY 440 PRCVESTAYCGPPPSINNGDTTSFPLSVYPPGSTVYRQCSFYKLGQSVTVTCRNKQWSE 499  
Db 237 --CVSTAYCGPPPSINNGDTTSFPLSVYPPGSTVYRQCSFYKLGQSVTVTCRNKQWSE 294  
QY 500 PPRCLDPCVSENNMKNKIQLKWRNDGKLYAKTGDAVEFQCKFPHKAMISSPPFRAICQ 559  
Db 295 PPRCLDPCVSENNMKNKIQLKWRNDGKLYAKTGDAVEFQCKFPHKAMISSPPFRAICQ 354  
QY 560 EGKFEYPICE 569  
Db 355 EGKFEYPICE 364  
RESULT 4  
ABG66700  
ID ABG66700 standard; Protein; 267 AA.  
XX AC ABG66700;  
XX DT 30-AUG-2002 (first entry)  
XX

DE XX Human novel polypeptide #35.

KW Human; inflammatory condition; shock; sepsis; immune response;

KW cancer; wound healing; central nervous system disease; haematopoiesis;

KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;

KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;

KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;

KW bone degenerative disorder; periodontal disease; reperfusion injury;

KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;

KW allergic condition; thrombolysis; thrombosis; coagulation disorder;

KW fungal infection.

XX Homo sapiens.

XX WO200244340-A2.

XX 06-JUN-2002.

XX 30-NOV-2001; 2001WO-US47004.

XX 30-NOV-2000; 2000US-0028952.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;

PI Yamazaki V, Ujwal ML, Drmanac RT;

XX WPI; 2002-508509/54.

DR N-PSDB; ABK94924.

XX Novel nucleic acids and polypeptides for diagnosis, treatment of

PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell

PT disorders, cancer and promoting wound healing -

XX Claim 10; Page 601; 672pp; English.

XX The invention relates to human novel polynucleotides and associated

CC polypeptides. The polynucleotides and polypeptides are useful for

CC treating inflammatory conditions such as arthritis, nephritis, Crohn's

CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses

CC and cancer and for promoting wound healing. The sequences are used to

CC induce the proliferation of neural cells and regeneration of nerve and

CC brain tissue, and are useful for the treatment of central and peripheral

CC nervous system diseases and neuropathies, such as Alzheimer's disease,

CC Parkinson's disease, Huntington's disease and amyotrophic lateral

CC sclerosis. The sequences are involved in chemotactic or chemokinetic

CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid

CC cell disorders and platelet disorders such as thrombocytopenia, tissue

CC regeneration of bone, cartilage, tendon, ligament and/or nerve, tissue

CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of

CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal

CC disease. The sequences of the invention are also useful for gut

CC protection or regeneration and treatment of lung or liver fibrosis,

CC reperfusion injury in various tissues, immune deficiencies and disorders

CC including severe combined immunodeficiency (SCID), bacterial or fungal

CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia

CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis

CC and coagulation disorders. Sequences ABG66666-ABG66758 represent human

CC novel polypeptides of the invention.

XX Sequence 267 AA;

Query Match 46.9%; Score 1492; DB 23; Length 267;

Best Local Similarity 100.0%; Pred. No. 9.2e-112;

Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 MIGNNMITCINGLWTELPWCVATHQKRCCKIAGVNIKTLLKSGKEFNHNSRIRYCSDI 362

DB 1 MIGNNMITCINGLWTELPWCVATHQKRCCKIAGVNIKTLLKSGKEFNHNSRIRYCSDI 60

QY 363 FRYRHSVCINGKWNPEVDCTEKREOFCPPPPPIPNQNMNTTNTVYODGKVAVLCKENYL 422

DB 61 FRYRHSVCINGKWNPEVDCTEKREOFCPPPPPIPNQNMNTTNTVYODGKVAVLCKENYL 120

QY 423 LPEAKEIVCKGRWQSLPRCVESTAYCGPPPSINNGDTTSFPLSVYPPGSTVYRCOSFY 482

DB 121 LPEAKEIVCKGRWQSLPRCVESTAYCGPPPSINNGDTTSFPLSVYPPGSTVYRCOSFY 180

QY 483 KLGCSVTVTCTRNKQWSEPPRCCLDPCVYVSEENMKNNTQLKWRNDGKLYAKTGDAVEPQCK 542

DB 181 KLGCSVTVTCTRNKQWSEPPRCCLDPCVYVSEENMKNNTQLKWRNDGKLYAKTGDAVEPQCK 240

QY 543 FPHKAMTSSPPFRAICQEGKFEYPICE 569

DB 241 FPHKAMTSSPPFRAICQEGKFEYPICE 267

RESULT 5

ABP41931

ID ABP41931 standard; Protein; 678 AA.

XX AC ABP41931;

XX DT 22-AUG-2002 (first entry)

XX DE Human ovarian antigen HVVBI38, SEQ ID NO:3063.

XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;

KW ovarian cancer; breast cancer; tumour; reproductive system disorder;

KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;

KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;

KW inflammatory condition; immune disorder; blood disorder;

KW cardiovascular disorder; respiratory disorder; neurological disorder;

KW gastrointestinal disorder; urinary system disorder; drug screening;

KW gene therapy; chromosome mapping; forensic analysis;

KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;

KW antiinflammatory; gynaecological; reproductive.

XX Homo sapiens.

OS WO200200677-A1.

XX 03-JAN-2002.

XX 07-JUN-2001; 2001WO-US18569.

XX 07-JUN-2000; 2000US-209467P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Birse CE, Rosen CA;

XX WPI; 2002-147878/19.

DR N-PSDB; ABQ55008.

XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,

PT useful in the prevention, treatment and diagnosis of cancer (e.g.

PT ovarian cancer), immune disorders, cardiovascular disorders and

PT neurological diseases -

XX Claim 11; SEQ ID No 3063; 2922pp; English.

XX The invention relates to 2175 novel human ovarian antigens (ABP41054-ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also

CC encompasses polypeptides 90% identical and polynucleotides 95% identical

CC to the sequences of the invention. The invention additionally relates to

CC recombinant vectors and host cells comprising human ovarian antigen

CC polynucleotides, antibodies against human ovarian antigens, and the use

CC of ovarian antigen polynucleotides and polypeptides in diagnosing,

CC treating, prognosing or preventing various ovary and/or breast-related

CC disorders. Such conditions include ovarian cancer and breast cancer, and

CC metastatic tumours of ovarian or breast origin, reproductive system

CC disorders (e.g., infertility, disorders of pregnancy, anovulation,

CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine

CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic

CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and



QY 61 SPKSPWTRITCTEESHPTPKCLRMCSFFVKNHSESSGLIHLEGGDTVOITCNTGYSL 120  
 Db 73 SPKSPWTRITCTEESHPTPKCLRLCFFPVENHSESSGQTHLEGGDTVOITCNTGYRL 132  
 QY 121 QNNEKNISCVBERGWSPTKSGECHVPILEAVDAOPKPKESYKVGDLKFSCKRNLI 180  
 Db 133 QNNEKNISCVBERGWSPTPKCRSTD----- 156  
 QY 181 RVGSDSVOCYOFOWSPNFPCTKGVRSCGPPPLNSGEGVKEIRKEEYGHNEVVEYDCNPN 240  
 Db 157 -----TSCVNPPTVQNAIXSRMSKYPSGGERVY----- 186  
 QY 241 FIINGPKKIQCVGENTLTCTCEQVKTCTGPIPELEYGYVQSPVPYHQHGVSEVNECRNE 300  
 Db 187 -----XCRSP 191  
 QY 301 YAMIGNMIFTCINGIWTPELPMCVATHQKRCCKIAGVNIKTLLKLSGKEFNHNSRIYRCS 360  
 Db 192 YEMFGDEVMCLNGNWE----- 209  
 QY 361 DIFRYHSVCINGKNWNEVDCTEKREQFCPPPPQIPNAQNMTTTTVYQDGKAVLCKEN 420  
 Db 210 -----PPQ----- 212  
 QY 421 YLLPEAKEIVCDRWQSLPRCVESTAYCGPPPSINNGDTTSFPLSVYPPGTVYTRCOS 480  
 Db 213 -----STGCKGPPPIDNGDITSFPLSVIAPASSVEYQCON 251  
 QY 481 FYKLGSSVTVTCRNKQWSEPPRCCLDCVSEENKNNIOLKWRNDGKLYAKTGDAVEFQ 540  
 Db 252 LYQLEGKRTICRNGOWSEPPKCLHPCVTSREIMENYIALRTAKLYXRTGESXEFV 311  
 QY 541 CKPPHKAMTSSPPFRAICQBGKEFYPTC 568  
 Db 312 CKRGYLRSSRSLRTTCMDGKLEYPTC 339

## RESULT 7

AA09065  
 ID AA09065 standard; Protein; 578 AA.

XX  
 AC AA09065;

XX  
 DT 06-JUL-1999 (first entry)

XX  
 DE Human complement factor H homolog protein.

XX  
 KW Human complement factor H; immunological mechanism; complement reaction;  
 KW gene therapy; immune stimulation; haematopoiesis regulation; chemotactic;  
 KW tissue growth activity; anti-inflammatory; tumour inhibition;  
 KW secretory signal.

XX  
 OS Homo sapiens.

XX  
 PN W09918200-A1.

XX  
 PD 15-APR-1999.

XX  
 PF 02-OCT-1998; 98WO-JP04448.

XX  
 PR 06-OCT-1997; 97JP-0272837.

XX  
 PA (PROT-) PROTEGENE INC.

XX  
 PA (SAGA ) SAGAMI CHEM RES CENT.

XX  
 PI Kato S, Sekine S;

XX  
 DR WPI; 1999-264019/22.

XX  
 DR N-PSDB; AAX34737.

XX  
 PT Human proteins with secretory signal sequences and nucleotide  
 PT sequences, useful in control of proliferation and differentiation of  
 PT cells

XX

PS Claims 1; Page 55-58; 71pp; English.

XX

CC This DNA encodes a protein having homology to human complement factor H,  
 CC which plays a role in the immunological mechanism involving the  
 CC complement reaction. The protein can also be used as an antigen for  
 CC preparing antibodies against the protein. The cDNA can be used as a probe  
 CC for gene diagnosis and the gene for gene therapy, as well as for large-  
 CC scale expression of the protein. The protein may also have immune  
 CC stimulating or suppressing activity, haematopoiesis regulating activity,  
 CC tissue growth activity, activin/inhibin activity, anti-inflammatory  
 CC activity, tumour inhibition activity, chemotactic/chemokinetic activity,  
 CC receptor/ligand activity, etc. The protein is identified by the presence  
 CC of a hydrophobic N-terminal secretory signal region, knowledge of the  
 CC protein function is not required, as in e.g. methods of expression  
 CC cloning.

XX  
 SQ Sequence 578 AA;

Query Match 31.5%; Score 1001.5; DB 20; Length 578;  
 Best Local Similarity 36.7%; Pred. No. 7.1e-72;  
 Matches 219; Conservative 94; Mismatches 238; Indels 45; Gaps 19;

QY 1 MLLFSVILLSWSTVGG-BGTLCDPPKIHGFLYDEEDYNPFSQVPTGTVFYSEYNF 59  
 Db 1 MLLLINVILTLWYSCANGOEKPCDFPEIQHGLGYKSLRRLYFPAAAGOSYSYCDQNF 60  
 QY 60 VPSKSFWRITCTEBSWPTPKCLRMCSPPFV--KNHSESSGLIHLEGGDTVOITCNTG 117  
 Db 61 VTFSGSYWDYIHTQDGNSTVPCLRTCKSDVEIEINGFISESSYIILNEETQYNCKPG 120  
 QY 118 YSLQ--NNEKNISCVBERGWSPTPIC-SFTKGECHVPILEAVDAOPKPKESYKVGDLKFS 174  
 Db 121 YATADGNSSGINTCLONGWSTQPIKFIK-----CDMPVEF-NSRAKNGMFKLHDTLDYE 175  
 QY 175 C---RKNLIRVSDSVQCQFGWSPNPTCKGQVRSCGPPPLNSGEGVKEIRKEEYGHNE 231  
 Db 176 CYDGYESYGNNTTDSIVCGEDGWS-HLPTCYNSESSECGPPPTISNGDTTSFPQKVLPS 234  
 QY 232 VREYDCNPNFIINGPKKIQCVGENTLTCTCEQVKTCTGPIPELEYG--YVQPSVPPY-- 287  
 Db 235 RVEYQCSYVELOGSKYVTCNGDNSEPPRCI-SMKPCEF-PEIQHGLHYENTRRPYFP 292  
 QY 288 -QHGVSVEYVNCNEYAMIGN---NMITCINGIWTPELPMCVATHQKRCCKIAGVNIKT-LL 342  
 Db 293 VATGQSYSYVCDQNFVTPSGSYWDYIHTQDGNLPTVPC-----LRTCKSDIEINGFI 347  
 QY 343 KLSGKEFNHNSRIYRCSDFRYRHS-----VCINGKNVPEVDCTEKREQFCPPPPQIP 396  
 Db 348 SESSSYIILNKETQYKCKPGYATADGNSSGSIITCLONGWSAQPICI----KFC-DMPVFE 402  
 QY 397 NQNMNTTNYODGKRVAVLCKENLLPEAK---EIVCKDGRWQSLPRCVESTAYCGPPP 453  
 Db 403 NSRAKNGMRPKLHDTLDYECYDGYEISYGNVTGSIYVCGEDGWSHPTCTYNSSEKCGPPP 462  
 QY 454 SINNGDTTSFPLSVYPPGTVYTRCOSYKLGSSVTVTCRNKQWSEPPRCCLDCPVVSEEN 513  
 Db 463 PISNGDTTSFLLKVVYVPSRVEYQCOSYVELOGSNYVTCNSGSEWSEPPRCIHPCIITEEN 522  
 QY 514 MNKNNIQLKWRNDGKLYAKTGDAVEPOCKFPHKAMLISSPPFRAICQBGKEFYPTCE 569  
 Db 523 MNKNNIQLKGRKSDIKYYAKTGDTIEPMCKLGYNANTSVLSFOAVCREGIVEYPRCE 578

## RESULT 8

AA082921

ID AAP82921 standard; protein; 660 AA.

XX

AC AAP82921;

XX

DT 26-NOV-1990 (first entry)

XX

DE B subunit of human Factor XIII.







Db	99	-----RDTSCVN-----PPTVQNA 111
Qy	399	QNMTTWN-YODGEKVAVLCKENYLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINN 457
Db	113	YIVSRQMSKYPSGERVRYQCRSPYEMFGDEEVMCLNGNWTBPQCKDSTGKCGPPPPIDN 172
Qy	458	GDTTSFPLSVYPPGSTVYTRCQSFYKLGVSVTVTCRNKQWSEPPRCLDPCVYSENM 514
Db	173	GDITSFPLSVYAPASSVEYQCNLYQLEGNKRITCRNGQWSEPPRCLHPCVISREIM 229
RESULT 12		
AAW39159		AAW39159 standard; Protein; 196 AA.
XX	AC	AAW39159;
XX	AC	AAW39159;
DT	27-APR-1998	(first entry)
XX	XX	Clone PRBS3FH2910 #4.1 CFH related protein fragment.
DE	XX	Complement factor H; tumour associated antigen; renal cancer;
XX	XX	urogenital cancer; medicament; modulator.
KW	XX	Synthetic.
XX	XX	W09738136-A1.
PN	XX	16-OCT-1997.
PD	XX	09-APR-1997; 97WO-US05710.
FF	XX	06-MAR-1997; 97US-0812481.
PR	XX	09-APR-1996; 96US-0015083.
PR	XX	09-APR-1996; 96US-0630048.
PR	XX	06-MAR-1997; 97US-0038614.
XX	XX	(BARD-) BARD DIAGNOSTIC SCI INC.
PA	XX	Enfield DL, Hass GM, Kinders RJ;
PI	XX	WPI; 1997-512742/47.
DR	XX	N-PSDB; AAV02795.
DR	XX	Treating or screening for cancer, e.g. renal or urogenital cancer -
PT	XX	by modulating or detecting tumour associated human complement Factor
PT	XX	H related antigen, or nucleic acid encoding it
XX	XX	Example 6B; Fig 7B; 104pp; English.
PS	XX	This partial protein is found in clone PRBS3FH2910 #4.1 and represents a
CC	XX	complement factor H related protein with homology to a region of the
CC	XX	human tumour-associated complement factor H (CFH). The detection of this
CC	XX	protein and a CFH antigen can be used in screening or for the treatment
CC	XX	of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.
CC	XX	Agents that may modulate this antigen could be used in the manufacture of
CC	XX	a medicament for the treatment of a tumour cell.
XX	XX	Sequence 196 AA;
SQ	XX	Query Match 13.6%; Score 432.5; DB 18; Length 196;
		Best Local Similarity 35.3%; Pred. No. 9.3e-27;
		Matches 84; Conservative 30; Mismatches 61; Indels 63; Gaps 3
Qy	287	YOHGVSVEVNCRNEYAMTGNMTCINGIWTPELPMCVATHQLKCKIACVNIKTLKLSG 346
Db	17	YKAGEPVTYTCATYKMDGASNVTCINSRWTRGPTC-----52
Qy	347	KEFNHNSRIRYRCSDIFRYRHSVCINGKWNPEVDCTEKREQFCPPPPQIPNAQNMTTVN 406
Db	53	-----RDTSCVN-----PPTVQNAIYVSROMS 74
Qy	407	-YODGEKVAVLCKENYLLPEAKEIYCKDGRWQSLPRCVSTAYCGPPPSINGDTSFPL 465

Db 75 KPSGERVRYQCSFYDMFDEEYMLCNGNWTETPPQCKDSTGKCGPPPPIDNGDITSFPL 134  
Qy 466 SVYPPGTVYRQSFYKLOGSVYTCRNKQWSEPPRCLDPCVYVSEENMKNNTQK 523  
Db 135 SVYAPASSVEYQCNLYQLEGKRNKRTCRNQWSEPPKCLHPCVLSREIMENYIALRW 192

RESULT 13  
AAR11139  
ID AAR11139 standard; Protein; 1087 AA.

XX AC AAR11139;  
XX DT 23-MAY-1991 (first entry)  
XX DE B lymphocyte membrane glycoprotein CR2.  
XX KW CR2; B lymphocyte; membrane receptor protein; Epstein-Barr virus;  
XX CC extracellular domain.  
XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
FT Region 20..90  
FT /note= "1"  
FT Region 20..153  
FT /note= "2"  
FT Region 20..276  
FT /note= "3"  
FT Region 20..652  
FT /note= "4"  
FT Region 20..776  
FT /note= "5"  
FT Region 20..1025  
FT /note= "6"  
FT Peptide 1..20  
FT Protein 21..1087  
FT /label= sig\_peptide  
FT /label= mat\_protein

XX MO9102536-A.  
XX 07-MAR-1991.  
XX 22-AUG-1990; 90MO-US04817.  
XX 23-AUG-1989; 89US-0398224.  
XX (SCRI-) SCRIPPS CLINIC & RE.

XX Moore MD, Cooper ND, Nemerow GR;  
XX WPI; 1991-087109/12.  
XX N-PSDB; AAQ10989.

XX Synthetic polypeptide(s) of extracellular domain - treat  
XX Epstein-Barr virus infections and diagnose same by formation of  
XX complex between CR2 and EBV

XX Disclosure; Fig 9.1-9.9; 60pp; English.

XX The six indicated fragments encode claimed and disclosed  
XX polypeptides, which are synthesised by recombinant expression, pref.  
XX in a baculovirus expression system in which a DNA plasmid, contg. a  
XX cDNA encoding this sequence, is truncated to encode a CR2 polypeptide  
XX that comprises a region of the extracellular domain of CR2. The  
XX construct is then inserted downstream of an appropriate promoter in  
XX a transfer vector and integrated into a baculovirus which is then  
XX used to infect host insect cells for expression.  
XX The polypeptides correspond to B-lymphocyte membrane receptor  
XX protein for Epstein-Barr Virus (EBV) and give specific binding.  
XX Protein number 3 is used to inhibit infection of mammalian cells in

CC contact with an aq. medium, (esp. mammalian blood) and of human  
CC B-lymphocytes, by EBV. The peptides are used to detect the presence  
CC of EBV in an aq. sample and to detect antibodies directed against CR2.  
CC The peptides are used in a pharmacological compsn. as active  
CC ingredient with a carrier to treat immune disorders.  
XX  
SQ Sequence 1087 AA:

Query Match 12.8%; Score 408; DB 12; Length 1087;  
Best Local Similarity 21.9%; Pred. No. 8.7e-24;  
Matches 170; Conservative 87; Mismatches 246; Indels 272; Gaps 38;  
Qy 26 PKIHGFLYDEEDYNPFQVPTGEVYFYSCEYFNVSFSPKSFWRITCTEFG-WSPTP--- 81  
Db 158 PMIHNGHTSEN---VGSFAPGLSVTYSCESGYLLVGEKI---INCLSSGKKSVAVPTC 210  
Qy 82 ---KCLRMCSPFPVKNGHSESSGLIHLEGTQVLIICNTGYSLONNEKNISCVERG---W 134  
Db 211 EEARCKSLGRFP---NGKVKEPPIIRV-GVTANFFCDEGYRLQGPSS-RCVIAGOGVAW 265  
Qy 135 STPPICSTKGECHVPIL-ANV-----DAOP----- 160  
Db 266 TKMPVCEETPCSPPPILNGRHIGNSLANVSYSVITYTCDPDEGVNELLIGESTLRC 325  
Qy 161 -----KKEYKYGDVDEKFSCKRNLIIR 181  
Db 326 TVDSQKTGWSGAPRCELSTSAVQCPHPQILRGMVSGQKDRYTYNDTVIFACMFGLT 385  
Qy 182 VGSDSVQCYQFG-WSNFPCTCKQVRSCGPPPLSGNGEYKEIRKEEYGHNEVVEYDCNP 240  
Db 386 KGSQKIQIRNAQGTWEPSPVCE---KECQAPPNILNGKEDRHMRVDFDGTSTIKYSCNPG 442  
Qy 241 FIINGPKTKQCV-DGEWT-TLPTCVOEVKTC-----GYPELEYGVVQPSV----- 284  
Db 443 YVLVGEESIQCTSEGVTTPVPOC--KVAACEATGRQLLTQKPOHOFVRDVSNSCGEGYK 500  
Qy 285 -----PP-----YQHGVSVEVNCRN--- 299  
Db 501 LSGSVYQECQGTIPWFMELRCKEITCPPPPVIYNGAHTGSSLEDFPYGTYTTCNPGP 560  
Qy 300 ---EYAMIGNMITCIN-----GIWT-ELPMCVATHOLKRCIAGVNIKTLLKLSKE- 348  
Db 561 ERGVFSLIGESTIRCTSNDOERGTSWGPAPICKLSLLAVQC--SHVHTANGYKISGKEA 618  
Qy 349 -FNHNSRIRYRGSDIFRYRHSVCINGK----WNPEVDCTEKRFQCPPPPOIPNAQNM-T 402  
Db 619 PYFYNDTVTFKCYSGFTLKSSQIRCKRDNTWDPPIPVCEKG--COPPPGLHGHRTGG 675  
Qy 403 TTVNYODGKAVLCKENYLLPEAKEIVC-KDGRWO-SLPRCVES----- 445  
Db 676 NTVFYVSGMTVDYTCDPGYLLVGNKSIHCMPSGNMSPSAPRCBETCQHVRSQSLQELPAGS 735  
Qy 446 -----TAYCGPPPSINNGDTTFPLS 466  
Db 736 RVELVNTSCODGYQLTGHAQYMCODAEENGIMFKKIPLCVKVICHCHPPVIVNGKHTGMABE 795  
Qy 467 VYPPGSTWYRQ-OSFYKL-QGSVTVTCRNKQW---SEPPRCL-----DPCVVYSEENMKN 516  
Db 796 NELYNGEVSVECDQGYLLGKNCSEVILKAWILERAFOCLRLSLCPNPEVKGHYKLNK 855  
Qy 517 -----NNIQLKWRNDGKLYAKTGDVAVEFOCKFPH-----KAMISSPP 553  
Db 856 THSAYSHNDIVVDCNPG--FIMNGSRV-IRCHTDNTWPGVPTCIKKAFICGPP 907

RESULT 14  
AAG66398  
ID AAG66398 standard; Protein; 979 AA.  
XX  
XX AAG66398;  
XX AC  
XX 16-OCT-2001 (first entry)  
XX

DE Receptor 222.  
KW Receptor 222; cytostatic; anti-HIV; cancer; HIV infection.  
XX Unidentified.  
OS CN1296952-A.  
PN CN1296952-A.  
XX 30-MAY-2001.  
XX 23-NOV-1999; 99CN-0124086.  
XX 23-NOV-1999; 99CN-0124086.  
XX (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.  
XX Mao Y, Xie Y;  
XX WPI; 2001-483894/53.  
DR N-PSDB; AAH75787.  
XX New polypeptide-complement receptor 222 for treating diseases, such as,  
PT cancer and human immunodeficiency virus infection -  
XX Claim 1; Pages 21-23 (Disclosure); 30pp; Chinese.  
XX The present sequence is the protein sequence for receptor 222. Receptor  
CC 222 and its coding sequence are useful in treating diseases e.g. cancer  
CC and HIV infection.  
XX Sequence 979 AA;  
Query Match 12.4%; Score 395; DB 22; Length 979;  
Best Local Similarity 23.2%; Pred. No. 8.4e-23;  
Matches 159; Conservative 92; Mismatches 242; Indels 192; Gaps 39;  
QY 23 CDFPKI-HHGFLYDEEDYNFSPQVPTGEVFYSCYENFVSPKSFWRITCTEAGS-PT 80  
DB 125 CDLPAPENGFL-----RTETSMGSAVQYCKPGHILAGSDDL--RLCLENRKWSGAS 175  
QY 81 PKCLRM-CSPPF-VKNGHSESSGLIHLEGDVVOITCNTGYSLQNNKNSICVERGW-STP 137  
DB 176 PRCEAISCCKPNPVMNGSIKSNLYL--STLYECDPGYVLTGRTTERRTCQDDKNWDEDE 233  
QY 138 PICSTKGECHVPILLEANVDAQPKESYKVGDLKFSCKRKNLIRVGSQSVQCYQFG-WSP 196  
DB 234 PIC--IPVDCSPPPVSAN--QGVGDEYTFQKEIYTCNEGFLLEGARSRVCLANGWSG 289  
QY 197 NFPTCKGQVRSCGPPPLNSGEVKEIRKEEYGHNEVVEYDCNPNFIINGPKKIQC-VDGE 255  
DB 290 ATPDCV-PVR-CATPQLANGVTEGL---DYGFMKEVTFHCHEGYLHGAPKLTCQSDGN 344  
QY 256 W-TTLPTCEQVKTCCYIPELEYGVQVSPYPPYQGVSVENCRNEYAMIGNNMITCI-N 313  
DB 345 WDAEIPLC--KPVNCGPEDLAHGF--PNGSFHGHIOYOCFFPGYKHLGNSRRCLSN 400  
QY 314 GIWT-ELPMCVATHQLKRCIKIAGVNIKTLKLSGKEFNHNSRIYRCSDFIRY----- 365  
DB 401 GWSGSSSCSLPC-----RCSFPIVIEYGI---VNGTDFDGCRAARQICFKGKLCLSEIT 453  
QY 366 -----RHSYC-----INCKW-----NPEVDCT 382  
DB 454 CEADQWSSGPHCEHTSCGSLPMIPNFISETSSKENVITYSCRCYVIOGSSDLICT 513  
QY 393 EK-----REQFCPPPIQIPNA----- 398  
DB 514 EKVMSQYPVCEPLSCGSPPSVANAVATGEAHTYSEVKLRCLBEGYTMDDTDTFTCOK 573  
QY 399 -----QNMFTTVNYOD---GEKVAVLCKENYLLPEAKEIYCK-DG 434  
DB 574 DGRWPPERISCPKCPIDENITHILVHGDDFVSNRQVSVSCABGYTTEGVNISVCOLDG 633  
QY 435 RMQ-----SLPRCVESTAYCGPPPSINNGDTSFPLSVYPPGSTVTVYRCQSFYKQLQGSVTVT 491

DB 634 TWEPFSDSC--SPVSCGKPESPHGFVVG---SKYTFESTIYQCPEGYELEGNERV 688  
QY 492 CR-NKOWS-----EPPRCLDPCVYSEENMN-KNNIQLKWRNDGKLYAKTGDAVEFOCKE 543  
DB 689 COENRQWGGVAICKETRCETPL-----EFLNGKADIENR-----TTGPNVVYSC-- 733  
QY 544 PHKAMISPPFERAICQE-GKFEYPI 567  
DB 734 -NRGYSLGSPSEAHCTENGTSHPV 757  
RESULT 15  
ABP37987  
ID ABP37987 standard; Protein; 1066 AA.  
XX -  
AC ABP37987;  
XX 23-JUL-2002 (first entry)  
DT DT  
XX Human GS96663 protein.  
DE Human; chromosome 9q31-34; lipoprotein metabolism disorder;  
KW cholesterol transport disorder.  
XX Homo sapiens.  
OS  
XX WO200071710-A2.  
PN  
XX 30-NOV-2000.  
PD  
XX 25-MAY-2000; 2000WO-FR01426.  
PF  
XX 25-MAY-1999; 99FR-0006587.  
PR 16-JUN-1999; 99US-0139450.  
XX (AVET) AVENTIS PHARMA SA.  
PA  
XX Denefle P, Rosier-Montus M, Arnould-Reguigne I, Prades C, Clepet C;  
PI WPI; 2001-025161/03.  
DR N-PSDB; ABN93421.  
XX New nucleic acid derived from human chromosome 9, used e.g. for  
PT diagnosis and drug screening, derived from genes implicated in  
PT disorders of lipoprotein metabolism -  
XX Claim 22; Page 262-266; 269pp; French.  
PS The present sequence is the protein sequence encoded by a human gene from  
CC chromosome 9q31-34. This sequence is likely to be involved in diseases  
CC of plasmatic lipoprotein metabolism, e.g. the reverse transport of  
CC cholesterol.  
XX Sequence 1066 AA;  
Query Match 12.4%; Score 395; DB 22; Length 1066;  
Best Local Similarity 23.2%; Pred. No. 9.4e-23;  
Matches 159; Conservative 92; Mismatches 242; Indels 192; Gaps 39;  
QY 23 CDFPKI-HHGFLYDEEDYNFSPQVPTGEVFYSCYENFVSPKSFWRITCTEAGS-PT 80  
DB 212 CDLPAPENGFL-----RTETSMGSAVQYCKPGHILAGSDDL--RLCLENRKWSGAS 262  
QY 81 PKCLRM-CSPPF-VKNGHSESSGLIHLEGDVVOITCNTGYSLQNNKNSICVERGW-STP 137  
DB 263 PRCEAISCCKPNPVMNGSIKSNLYL--STLYECDPGYVLTGRTTERRTCQDDKNWDEDE 320  
QY 138 PICSTKGECHVPILLEANVDAQPKESYKVGDLKFSCKRKNLIRVGSQSVQCYQFG-WSP 196  
DB 321 PIC--IPVDCSPPPVSAN--QGVGDEYTFQKEIYTCNEGFLLEGARSRVCLANGWSG 376  
QY 197 NFPTCKGQVRSCGPPPLNSGEVKEIRKEEYGHNEVVEYDCNPNFIINGPKKIQC-VDGE 255

Db 377 ATPDCV-PVR-CATPQLANGVTEGL---DYGFMEKVTFFHCHGEGYILHGAPKLTQCSQDGN 431  
Qy 256 W-TTLPTCQVQKTCGYIPELEYQVOPSPYOHGYSVEVNCNREYAMIGNNMITCI-N 313  
Db 432 WDAEIPLC--KPVNCGPPEDLANGF--PNGFSFIHGHHIQYOCFPGYKLGHSRRCLSN 487  
Qy 314 GIWT-ELPMCVAHQKRCIAGVNIKTLLKSGKEFNHNSRIRYRCSDFRY----- 365  
Db 488 GSWGSGSPCLPC---RCSTPVEYGT---VNGTDFDCGAARIQCFKGLLGLSEIT 540  
Qy 366 -----RHSVC-----INGKW-----NPEVDCT 382  
Db 541 CEADGQWSGFPHCETSCGSLPMIPNAFISSETSMKENVITYSCRSYVIOGSSDLICT 600  
Qy 383 EK-----REQFCPPPPQIPNA----- 398  
Db 601 EKGWNSQPYPCPEPLSCGSPSPVANAVATGEAHTYSEVKRLCLEGYTMDTDTDTTCQK 660  
Qy 399 -----QNMTTVNYQD---GEKVAVLCKENYLLPEAKEIVCK-DG 434  
Db 661 DGRWFERISCSPPKCPLENITHILVHGDDFSVNRQVSVCABGYTFEGVNISVCOLDG 720  
Qy 435 RWQ---SLPRCVSTAYCGPPPSINNGDTSFPLSVYPPGSTVYRCQSPYKLGSVTVT 491  
Db 721 TWEPFESDESC--SPVSCGPSPSPHGVVVG---SKYTFESTIYQCEPGYEGLEGNRERY 775  
Qy 492 CR-NKQWS-----EPPRCLDPCVSENNN-KNNIQLKWRNDGKLYAKTGDAVEFOCKF 543  
Db 776 COENROWSGVAICKETRCETPL-----EFLNGKADIENR-----TTGPNVYVSC-- 820  
Qy 544 PHRAMISSPPFRAICQF-GKFEYPI 567  
Db 821 -NRGYSLEGPSEAHCTENGTSHPV 844

Search completed: July 17, 2003, 13:50:54  
Job time : 77 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 13:49:33 ; Search time 85 Seconds  
(without alignments)  
1379.304 Million cell updates/sec

Title: US-09-805-337A-2

Perfect score: 3183

Sequence: 1 MLLFSVILISWSTVGEG.....SPPFRAICQGRFEYPICE 569

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phage.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3183	100.0	569	4	Q9BXR6	Q9BXR6 homo sapien
2	1524	47.9	808	11	Q61408	Q61408 mus musculus
3	1463.5	46.0	1236	11	Q91YB6	Q91YB6 rattus norv
4	1426	44.8	699	11	Q91WX0	Q91WX0 rattus norv
5	1355.5	42.6	1172	4	Q9NU87	Q9NU87 homo sapien
6	1284	40.3	657	4	Q14006	Q14006 homo sapien
7	906.5	28.5	343	11	Q61406	Q61406 mus musculus
8	683.5	20.8	509	11	Q8R018	Q8R018 mus musculus
9	646	20.3	452	11	Q61407	Q61407 mus musculus
10	476	15.0	1053	13	Q91275	Q91275 paralabrax
11	435.5	13.7	1045	6	Q46545	Q46545 ovis aries
12	393.5	12.4	3567	11	Q9ES77	Q9ES77 mus musculus
13	387.5	12.2	303	11	Q61405	Q61405 mus musculus
14	387.5	12.2	1911	6	Q29528	Q29528 papio hamad
15	387.5	12.2	2014	6	Q29530	Q29530 pan troglod
16	383	12.0	2039	4	Q16745	Q16745 homo sapien

#### ALIGNMENTS

##### RESULT 1

Q9BXR6 ID Q9BXR6 PRELIMINARY; PRT; 569 AA.  
AC Q9BXR6;  
DT 01-JUN-2001 (TREMREL. 17, Created)  
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)  
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)  
DE Complement factor H-related protein 5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCHI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21125890; PubMed=11058592;  
RA McRae J.L., Cowan P.J., Power D.A., Mitchellhill K.I., Kemp B.E.,  
RA Morgan B.P., Murphy R.F.,  
RT "Human Factor H-related Protein 5 (FHR-5). A New Complement-Associated Protein.";  
RL J. Biol. Chem. 276:6747-6754(2001).  
DR EMBL; AF295327; AAK15619.1; .  
DR HSP; P10998; 1VVD.  
DR InterPro: IP000436; Sushi\_SCR\_CCP.  
DR Pfam; PF00084; sushi; 8.  
DR SMART; SM00032; CCP; 8.  
SQ SEQUENCE 569 AA; 64419 MW; 7FAAE31707B0C112 CRC64;

Query Match 100.0%; Score 3183; DB 4; Length 569;  
Best Local Similarity 100.0%; Pred. No. 9.5e-275; Indels 0; Gaps 0;  
Matches 569; Conservative 0; Mismatches 0;

Qy 1 MLLFSVILISWSTVGEGTLCDFPKIHGHGLYDEEDYNPFQVPTGEVFFVYSCYNFV 60

|||||

1 MLLFSVILISWSTVGEGTLCDFPKIHGHGLYDEEDYNPFQVPTGEVFFVYSCYNFV 60

|||||

61 SPKSFVTRITCTEGNSPTPKLRMCSFPVKNGHSESSGLIHLEGTVOICNTGYSL 120

|||||

61 SPKSFVTRITCTEGNSPTPKLRMCSFPVKNGHSESSGLIHLEGTVOICNTGYSL 120

```

QY 121 ONNEKNISCVRGHSTPPICSFTKGECHVPILEANVDAQPKKSKYKVDVLFKSCRNL 180
Db 121 ONNEKNISCVRGHSTPPICSFTKGECHVPILEANVDAQPKKSKYKVDVLFKSCRNL 180
QY 181 RVGSDSVQCYQFGWSPNFTCKGVRSCGPPPOLSGEVEIKRKEEYGHNEVEYDCNPN 240
Db 181 RVGSDSVQCYQFGWSPNFTCKGVRSCGPPPOLSGEVEIKRKEEYGHNEVEYDCNPN 240
QY 241 FIINGPKKIQCVGEMTLLPTCVQVKTGCIPELEYGYVQSPVPPYOHGVSVEVNCNE 300
Db 241 FIINGPKKIQCVGEMTLLPTCVQVKTGCIPELEYGYVQSPVPPYOHGVSVEVNCNE 300
QY 301 YAMIGNNMITCINGITELPMCVATHOLKCKTAGVNIKTLLKLSGKEFNHNSRIYRCS 360
Db 301 YAMIGNNMITCINGITELPMCVATHOLKCKTAGVNIKTLLKLSGKEFNHNSRIYRCS 360
QY 361 FIINGPKKIQCVGEMTLLPTCVQVKTGCIPELEYGYVQSPVPPYOHGVSVEVNCNE 420
Db 361 FIINGPKKIQCVGEMTLLPTCVQVKTGCIPELEYGYVQSPVPPYOHGVSVEVNCNE 420
QY 421 YLLPEAKEIKVCKDRMOSLPRVESTAYCGPPPSINNGDTSFPLSYPPGSTVYRCQS 480
Db 421 YLLPEAKEIKVCKDRMOSLPRVESTAYCGPPPSINNGDTSFPLSYPPGSTVYRCQS 480
QY 481 FYKLGQSVTTCRNKOWSEPPRCCLPCVYSEENMKNNTOLKWRNDGKLYAKTGDAVEFQ 540
Db 481 FYKLGQSVTTCRNKOWSEPPRCCLPCVYSEENMKNNTOLKWRNDGKLYAKTGDAVEFQ 540
QY 541 CKFPHKAMISSPPFRAICQEGKEFYPICE 569
Db 541 CKFPHKAMISSPPFRAICQEGKEFYPICE 569

```

## RESULT 2

```

Q61408 PRELIMINARY; PRT; 808 AA.
ID Q61408;
AC Q61408;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Complement factor H-related protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90153969; PubMed=1689298;
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA Chaplin D.D.;
RT Identification and sequence analysis of 4 complement factor H-related
RT transcripts in mouse liver.
RL J. Biol. Chem. 265:3193-3201(1990).
DR EMBL; M29009; AAA37416.1;
DR HSSP; P08603; 1HCC.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 13.
DR SMART; SM00032; CCP; 13.
SQ SEQUENCE 808 AA; 91654 MW; 6FD97D53CE74DF6D CRC64;

```

```

Query Match 47.9%; Score 1524; DB 11; Length 808;
Best Local Similarity 36.2%; Pred. No. 6.4e-127;
Matches 292; Conservative 87; Mismatches 187; Indels 240; Gaps 6;

```

```

QY 1 MLLFSVLISWSTVGGEGTLCDFPKIHGFLYDEEDYNPFSSQVPTGVEFYVCEYNFV 60
Db 6 MLLSNLLTAMLSATGKGVKSCERPQFKYGLYEEILRPFPVSYGNKYKCKDNGFS 65
QY 61 SPSKSFWRITTEGWSPTPCLRMCSPPFYKNHSGESSGIHLFGDTVQICNTGYSL 120
Db 66 PPSGLFWLDYIRCTVGQWKVEPCVRKCVFHYVVEGEFAFWKXIYVQGSQSLKVCYNGYSL 125
QY 121 QNNEKNISCVRGHSTPPICT-----SF----- 142

```

```

Db 126 QNGQDTMTCTENGSPPPKIRIKTCSVSDIEIENGFFSEFRTYALNRETYSRCKQYV 185
QY 143 ----- 142
Db 186 TNGTGRSRTCLONGWSPSCIKSCERPVSNTKNNSTWFKLNDKLDYECLIGHEN 245
QY 143 -----TKG-----ECHVPFILEANVDAQPKKESYKVDVLFKSCRNL 180
Db 246 EYKHTKGSITCTYGVGSDTSCPYEIECSVPIIDRLKLVSPRKEKYRVGDLLESCRS -H 304
QY 181 RVGSDSVQCYQFGWSPNFTCKGVRSCGPPPOLSGEVEIKRKEEYGHNEVEYDCNPN 240
Db 305 RVGSDSVQCYQFGWSPNFTCKGVRSCGPPPOLSGEVEIKRKEEYGHNEVEYDCNPN 240
QY 241 FIINGPKKIQCVGEMTLLPTCVQVKTGCIPELEYGYVQSPVPPYOHGVSVEVNCNE 300
Db 365 FLKGPKNKIQCVGEMTLLPTCVQVKTGCIPELEYGYVQSPVPPYOHGVSVEVNCNE 424
QY 301 YAMIGNNMITCINGITELPMCVATHOLKCKTAGVNIKTLLKLSGKEFNHNSRIYRCS 360
Db 425 FTMIGLSVSCLSGKWTQLPKCVATDLEKCRVLKSTDTIEAIKPRNEFOHNTMYKCR 484
QY 361 DIPRYHSHVCINGKWNPEVDCTEKREOFQCPPPPQIPNAQNMHTTVNYQGEKVAVLCKEN 420
Db 485 DKOEYEHSHVCINGKWNPEVDCTEKREOFQCPPPPQIPNAQNMHTTVNYQGEKVAVLCKEN 420
QY 421 YLLPEAKEIKVCKDRMOSLPRV----- 443
Db 543 YLTQDPEEMCKDRMOSLPHCIGLPCGPPPSILRGTVSLELESYOHGEEVTHCSTGFG 602
QY 444 ----- 443
Db 603 IDGPAFICEGGKWSDPKCIKTNCDVLTPIENAIIRGKKKSYRTGEQVTPRCQSPYOM 662
QY 444 -----ESTAYCGPPPSINNGDTSFPLSYPPGSTVYRCQSFPYK 483
Db 663 NGSDTVTCVNSRWIGKPVCKDSRGKCGPPPIDNGDITSLSLPEYFPSSVDYQCKYLL 722
QY 484 LOGSVTTCRNKOWSEPPRCCLPCVYSEENMKNNTOLKWRNDGKLYAKTGDAVEPOCF 543
Db 723 LAGKKTTCRNKOWSEPPRCCLPCVYSEENMKNNTOLKWRNDGKLYAKTGDAVEPOCF 543
QY 544 PHKAMISSPPFRAICQEGKEFYPICE 569
Db 783 GYHKARGSPFRTKCIISGTINPTCE 808

```

## RESULT 3

```

Q91YB6 PRELIMINARY; PRT; 1236 AA.
ID Q91YB6;
AC Q91YB6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Complement inhibitory factor H.
OS Rat.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RA Demberg T., Goetze O., Schlaf G.;
RT "Rat complement factor H: molecular cloning, sequencing and expression
RT in tissues and isolated cells.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ320522; CAC67513.1;
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00084; sushi; 20.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.

```

```

SQ  SEQUENCE 1236 AA; 140343 MW; 1AC89FFA28232EBF CRC64;
Query Match 46.0%; Score 1463.5; DB 11; Length 1236;
Best Local Similarity 31.7%; Pred. No. 2.6e-121;
Matches 289; Conservative 84; Mismatches 172; Indels 367; Gaps 4;

Qy 23 CDPKIHGFLYDEEDYNPSQVPTGEVFYISCEYNFVSPKSFWTITCTEGWSPTPK 82
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 325 CDPFQKHGLYDEESRPFPVPIGKEYSYCDNGFTTPSQSYWDYLRCTVNGWEPEVP 384
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 83 CLRMCSFPFKNGHSESSGLIHLEGDVQLICNTGYSLONNEKNISCEVERGWSPPIC-- 140
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 385 CLROCIHFHYEGESLWQRYIEGQAKVQCHSGSLPNQDQDITLCTENGWSPPKCVR 444
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 141 ----- 140
Db 445 IKTCVSVDIEIENGFFSESDYTYALNRKTRYCKQGYVTNTGETSGIITCLQDQWSRPS 504
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 141 -----SFTKG----- 145
Db 505 CIKSCDMPVFENAMTKNNWFKLNDKLDYECHEGYENYKHTKGSITCTYDQWSSPSC 564
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 146 ---ECHVPILLEANVDAOPKESYKVGDLAKFSCRNLIYVGSQVQCYQFQWSPFTCK 202
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 565 YERCSIPLRQDLVVPREVYKVGDSLFSRSG--HRVGADLVQCYHFQWSPFTCE 623
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 203 QVRSCGPPQLSNGEYKEIRKEEYGHNEVEYDCNPNFIINGPKKIQCYDGEWTTLPIC 262
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 624 GOVKSCDPLEIPNGEIKTKYKVEYSHGVDVEYDCKPRFLKGNKLCQYDQKWTTLPIC 683
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 263 VEQVKTCGYIPELYGVQVSPVYQHGVSVEVNCRNEYAMIGNMIMTICINGIWTLPIC 322
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 684 VEYERTCGDLPELHSGSVKSLVPPYHHGDSVEFTCTETFTWIGHAVFVCSGRWTELPOC 743
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 323 VATHOLKCKTAGVNIKTLLKSGKEFNHNSRIYRCSDFIRYHSHVCINGKNWPEVDC 382
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 744 VATDLKCKRKPSTGIDAIHPNKEFNHNSFVSIRCRQKQOEYHSHVCINGKNWPEVDC 803
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 383 EKRFQCPPPPQIPNAQNTMTTYQDGEKAVLCKENYLLPEAKEIVCKDGRWQSLPRC 442
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 804 RNEKRFQCPPPPQIPNAQVETTVKYLQGEKSVLCQDGYLTQGPPEVCKHGRWQSLPRC 863
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 443 VE----- 444
Db 864 TEKIPCSQPKIEHGSIKSPRSEERDLIESSSYEHGTTFTSYVCDGDFRISENRVTCNM 923
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 445 ----- 444
Db 924 GKWSLPRCVICPGPPPSIPLGLVSHLELSYQYGEVYVNCSEGNLIDGPAFIKCVGGQ 983
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 445 ----- 444
Db 984 WSEPPKCIKTDCLNLTPEIAKPTKPKKKYSRSGEQVTFRCPPPYRMDGSDIVTCVNTKW 1043
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 445 ----- 444
Db 1044 IGQPVCKDNKSCVNPVHPNATILTRHKTYPSCDKVRYDCNKPPELFWGSGDVPRKTFW 1103
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 445 -----STAYCGPPPSINNGDTTSPPLSVYPPGSTVTVRCQSFYKLOGSVTVTCRNKQ 496
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1104 TEPKCKDSTCKCGPPPIDNGDITSLPVPAPLSSVEYQCNYYLLKGNKIVTCRNKG 1163
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 497 WSEPPRCIDPCVSEENNNKNIQLKRWNDGKIYAKTGDAVEFOCKFPFKAMISSPPFRA 556
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1164 WSPPTCLHACVIPEDIMEXHNIVLRWENAKIYSQSENIETFFMCKPGYRFRKRGSPFRT 1223
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 557 ICQEGKFEYPTC 568
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1224 KCIEGHINPTC 1235
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

RESULT 4
Q91WX0
```

```

ID Q91WX0 PRELIMINARY; PRT; 699 AA.
AC Q91WX0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Complement factor H-related protein.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY;
RA Ren G.; Quigg R.J.;
RT "Rat complement factor H-related protein sequence.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF436847; AAL25802.2;
SQ SEQUENCE 699 AA; 78309 MW; 2CDC1B0D1C1FDA907 CRC64;

Query Match 44.8%; Score 1426; DB 11; Length 699;
Best Local Similarity 44.7%; Pred. No. 2.8e-118;
Matches 263; Conservative 76; Mismatches 174; Indels 75; Gaps 4;

Qy 53 YSCYENFVSPKSFWTITCTEGWSPTPKCLRMCSFPFVKNGHSESSGLIHLEGDVQI 112
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 113 YVCKQGYTTANGERSGSITCLQTGWSAQPSCKISCDVPVFENAKSKNDSTWFKLNDKLDY 172
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 113 ICNTGY--SLQNNKKNISCEVERGWSPTPICSTFGEGHVPTLEAVDAQPKESYKYGDV 170
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 173 ECHGYENKHYKTHGSIIMCTDQWSDKPS--YEICSIPTLDRPLVVPYPRKIYTVGDL 230
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 171 LKFSCKNLIRVGSQVQCYQFQWSPFPCTCKGQVRSCGPPQLSNGEYKEIRKEEYGHN 230
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 231 LKFSCKPG--HRVGDSVQCYDQFQWSPFPCTCKGQVRSCGQYPELLNGEIKTKRAEYSHG 289
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 231 EVYEDCNPNFIINGPKKIQCYDGEWTTLPICVOQVTCGVPIPELYGVQVSPVTPYQHG 290
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 290 DVVEYDCKPRFLKGNKIQCYDQKWTTLPTCBEKRTCTGDLPELHSGVSKFSPYHHG 349
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 291 VSVVNCRNEYAMIGNMIMTICINGIWTLPICVATHOLKCKIAGVNIKTLLKSGKEFN 350
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 350 DSVEFTCAEFTVLQGSVSCISGRWAQLPRCAVDQLEKCAKPLAVIGTNKLLKDGDLN 409
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 351 HNSRIYRCSDFIRYHSHVCINGKNWPEVDCTEKREQFCPPPPQIPNAQNTMTTVNYQDG 410
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 410 HNSISYKGGSQYGYGICINGRWDPEPTCTREKTFCPPPQIPNAHVETTVTKYLDG 469
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 411 EKAVLCKENYLLPEAKEIVCKDGRWQSL----- 439
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 470 EKSVLCODGYLTQGPPEVCKHGRWQSLRPTAKITCSQPKIDHGSIKLPLRSLDAGRD 529
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 440 -----PRCEVSTAYCGPPPSINNGDT 460
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 530 AIEYSSHEHGTTFYVCDGDFRISKENGVTCHMKGNSSPPRCVESTGCKGPPFTIENGDI 589
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 461 TSPPLSVYPPGSTVTVRCQSFYKLOGSVTVTCRNKQWSEPPRCIDPCVSEENNNKNIQ 520
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 590 TSLPLVPVPLSSVEYQCYFQWSPFKKIKTKNGEWEPEPKCLHACVLTBEIMRRONII 649
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy 521 LKRWNDGKIYAKTGDAVEFOCKFPFKAMISSPPFRAICQEGKFEYPTC 568
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 650 PKRTKNKIYIOSGDYVEFVCLTAYQAQGSPEFRTQCIDGHINPTC 697
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

RESULT 5
Q9NU87
ID Q9NU87 PRELIMINARY; PRT; 1172 AA.
AC Q9NU87;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DJ177P10.1.1 (H factor 1 (complement) isoform 1).
GN HFL.
```

OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Bird C.;
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AL049744; CAB70597.1; ..
DR	HSSP; P08603; 1HFH.
DR	InterPro: IPR000436; Sushi_SCR_CCP.
DR	Pfam: PF00084; sushi; 19.
DR	SMART; SM00032; CCP; 19.
SQ	SEQUENCE 1172 AA; 132087 MW; 8F5B954C4B4FA54 CRC64;
<p>Query Match 42.6%; Score 1355.5; DB 4; Length 1172;            Best Local Similarity 30.0%; Pred. No. 1e-111;            Matches 271; Conservative 80; Mismatches 136; Indels 417; Gaps 4;</p>	
Qy	23 CDFFKIIHHGFLYDEEDYNPFSSQVPTGEVYFYSCYINFPVSFKSFWTRITCTEEGWSPTPK 82
Db	325 IDYDPDIKHGGLYHENMRPYPVAVGKYYSYCDHEHFPETPSGSDWDHICTQDQMSPAVP 384
Qy	83 CLRWCSPPFYKNGHSESSGLIHLBGDFVQIICNTGYSLQNNKNIISCVERGWSPTPIC-- 140
Db	385 CLRCYCFPYLENGYQNHRGRFVGOKSDVACHFGYALPKAQTIVTCMENGSWSPTRCIR 444
Qy	141 -----SFTKG----- 145
Db	445 VKTCSKSIDIENGFISEQYTYALKKAKYOCKLVGYVADGETSGSITCGKQGSQAPT 504
Qy	146 ----- 145
Db	505 CIKSCDIPVFMNARTKNDTFWFLKNDLTLDYEDCHDYESNTGSTTGSIVCGYNGWSLDPIC 564
Qy	146 ---ECHVPILLEANYDAQPKESYKGVGYLKFSCRNILIRVGSDSVOCYQFGWSNFFTPCK 202
Db	565 YERECELPKIDVHLVDPORKKQDQYKGVGLKFSCKPGFTIVGPNVSQCYHFGLSPLDPICK 624
Qy	203 GQVRSCTPPQSLNGEYKIRKEYGHNVEVYDCNPNFIINGPKKIQCVDGEMTLPIC 262
Db	625 EOVSGCSPPELLNGVNEKTEKEYGHSBVEVYVCNPRFLMKGNPKIQCVDGEMTLPVC 684
Qy	263 VEQVKTGCIPELEYGYVQPSVPYOHGVSVEVNCRNEYAMIGNMIMTCINGIWTLPWC 322
Db	685 IVEBTECGDIPLEHGHGAQLSSPPYYGDSVEPNCSSEFTMIGHRSITCIHGVTQLPQC 744
Qy	323 VATHQLKRCIKAGVNIITLLKSGKENHNSIRYRCSDIPRYKHSVCINGKMNPEVDCT 382
Db	745 VVA----- 747
Qy	383 EKREOFCCPPPOIPNNAONMTTIVYQDGEKVAVLCKENYLLPEAKEIVCKDGRWOSLPRC 442
Db	748 --QIQLCPPPPQIPNNSHMTITLNYRQGEKVSVLQENYLIQEGEETICKDGRWOSLPLC 805
Qy	443 VE----- 444
Db	806 VEKIPCSQPQIEHGTINSSRSQESYAHGCTKLSYTCGGFRISENETTCYMKCKWSPP 865
Qy	445 ----- 444
Db	866 QCEGLPCKSPPEISHGVYAHMSDSYQYEEVYKCFEGFGIDGPAIAKCLGKWSHPPSC 925
Qy	445 ----- 444
Db	926 IKTDCLSLSPFENAIPMGEKKDVKYKAGEQVYTCATYKMDGASNVTCINRSWTGRPTCR 985
Qy	445 ----- 444
Db	986 DTSCVNPPTVQNAVIVSRQMSKYPGSRVRVYQCRSPYEMFGDEVMCLNGNWTPEPQCKD 1045
Qy	445 STAYCGPPPSINNGDTPSLSVYPCGSTVTVYRCQSFYKLGQSVTVTCRNKQNSPEPRCL 504

```

Db 1046 STGKCGPPPIDNGDITTSFFLSVYAPASSVEYQOONLYQLEGKRRITCRNGOWSEPPKCL 1105
Qy 505 DPCVVSSENNKNNIOLKWRNDCKLVAKTGDVAFEOFCKPPHKAMISSPPPRATCQSGKPE 564
Db 1106 HPCVISEIMEWNYNIALRWTAOKLISRTGESVEFVCKRGYRULSSRSHLTURTCWDGKLE 1165
Qy 565 YVIC 568
Db 1166 YPTC 1169

RESULT 6
Q14006 PRELIMINARY: PRT; 657 AA.
AC Q14006
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Complement H factor (Fragment).
GN HF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88025472; PubMed=2889480;
RA Day A.J., Ripoché J., Lyons A., McIntosh B., Harris T.J., Sim R.B.;
RT "Sequence analysis of a cDNA clone encoding the C-terminal end of
RT human complement factor H";
RL Blosci. Rep. 7:201-207(1987).
DR EMBL; M17517; AAAS2016.1; -.
DR HSSP; P08603; 1HFH.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00084; sushi; 11
DR PRINTS; PR00619; GATAZFINGER.
DR SMART; SM00032; CCP; 11.
FT NON_TER 1
SQ SEQUENCE 657 AA; 74247 MW; 6862CDG5239F1213 CRC64;

Query Match 40.3%; Score 1284; DB 4; Length 657;
Best Local Similarity 38.1%; Pred. No 1.1e-105;
Matches 246; Conservative 59; Mismatches 103; Indels 238; Gaps 1;

Qy 161 KKEYKVGVDLVKFSCKNLIRVSDSVQCYQFGWSPNFTCKGVRSCGPPPLSGEYK 220
Db 9 KKDQYKVGVEVKFSCKPGFTVPSVQCYHFGLSPLDPIKREYQVSGCPPPELLNGNVK 68
Qy 221 EIRKEYGHNVEYDCNPNFINGPKKTCQVDGWTTLPTCUEQVKTGCTGYDELYGVV 280
Db 69 ERTKEYGHSEVVEYCYNPRFLMKGNPKIQCYDGEWTTLPVCIVEESTCGDIPLEHGW 128
Qy 281 QPSVPPYQGVSVVEVNCREYAMGNMTTCINGIWTELPKCVATHOLKRCIKAGVNIKT 340
Db 129 QLSPPYYGDSVEFNCSFSFMIGHRSITCHGVWVTLQPCVADKLKCKSSNLTILE 188
Qy 341 LLKLSKERNHNSIRIRYCSDFIRYHVSVCINGKNKNEPVDCTEKREQFCPPPPQIPNAQN 400
Db 189 EHLKNKEFDHNSIRIRYCRKGEGWTHVTCINGRNDPEVNCSSMAQIQLCPPPPQIPNSHN 248
Qy 401 MTTVNYQDGEKVAVLCKENYLLPEAKETVCKDGRWQSLPRCWE----- 444
Db 249 MTTLLNYRDEKGVSVLCQENYLIQEGEETICKDGRWQSLPRCWEIKPSQPPQIEHGTIN 308
Qy 445 ----- 444
Db 309 SSRSSQESYAHGTKLSYTCGEGFRISEENETTCYMGKNSSPPQCBLPKCKSPPEISHGVV 368
Qy 445 ----- 444
Db 369 AHMSDSYQYGEVTVKCFBGFGIDGPATAKCLGKWSHPPSPCKIKTDCLSLSPFENAI PMG 428

```





```

Db 343 WFKLNDKLDYBCLVGFENEXKHTKSGITCTYVGSWSDTSCYDSTRTGCGPPPIDNGDITS 402
QY 463 FPLSYVPPGSTVTVRCQSFYKLOGSVTVTCRNKQWSEPPRCCLDPCYVSEENKNNIOLK 522
Db 403 LSLPEYELSSVDYOCQYIILKGNKTITCRNGKWSSEPTCLHACVIPENIMEAHNIILK 462
QY 523 WRNDGKLYAKTDAVEFOCKEPHKAMISSPPFRAICQEGKFEYPTC 568
Db 463 WRHTEKIYHSGEDIEPECKRGYKARGSLPFRKTCINGTINYPTC 508

RESULT 9
Q61407
ID Q61407 PRELIMINARY; PRT; 452 AA.
AC Q61407;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Complement factor H-related protein (Fragment).
GN A1194696.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA Chaplin D.D.;
RL Submitted (APR-1990) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90153969; PubMed=1689298;
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA Chaplin D.D.;
RT "Identification and sequence analysis of 4 complement factor H-related
RT transcripts in mouse liver."
RL J. Biol. Chem. 265:3193-3201(1990).
DR EMBL; M29010; AAA37415.1;
DR HSSP; P08603; 1HFH
DR MGD; MGI:2138159; A1194696.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 7.
DR SMART; SM00032; CCP; 7.
FT NON_TER 452 452
SQ SEQUENCE 452 AA; 51602 MW; 2B697A4FFC6E13CA CRC64;

Query Match 20.3%; Score 646; DB 11; Length 452;
Best Local Similarity 24.9%; Pred. No. 4.3e-49;
Matches 157; Conservative 66; Mismatches 160; Indels 248; Gaps 13;

QY 1 MLLFVILLISVSTVGGEG----- 20
Db 6 MLLLSMILLTAWLSATKGEKTCSPYILNGIYTPHRIHRSDEIRYECNYGFVPTGS 65
QY 21 -----TL--CDPPKIHGFLYDEEDYNPFQSVPTGEVYISCEYNFVSP 62
Db 66 TVSKCTPTGWIPVPRCTLAPCEPPQKRYGLYEEISLRNFPVIGNKYSYRCNDGFSPP 125
QY 63 SKSFWRITCTEGWSPKRLMCSFPFVKNGHSSGLIHLEGDTVQIICNTGYSLQN 122
Db 126 SGYSMDYLRLCTAGWEPEVPCVKCFVHYVNGDSAYWEKIVQGSGLKVCYNGYSLQN 185
QY 123 NEKNISCVRGWSTPDCISFTKGECVCHVPILEANVDAOPKESYKVGDLVKFSCRMLIRV 182
Db 186 GDIMTCTENGWSPPPK-----TRI 206
QY 183 GSDSVQCYOGWSPNFPPTCKGQVRCGPPPLNSGEVKEIRKEEYGHNEVVEYDNCNPTI 242
Db 207 -----KTCAS-----DIHIDNGLFLSP--SYIYALNRETSTYRCQGVY 243
QY 243 INGPK---KIQCVDGWTPLTFCVBOVKTGCIPELEYGVQVSPPPYGHGVSEVNCRN 299
Db 244 TNTGTSIGSITCLQNGWSPQSC---IKSC-----DMPVFENSIT----- 280

```

```

QY 300 EYAMIGNMITCINGIMTELPVCVATHQKCRCKIAGVNIKTLLKLSGKEFNHNSRIRYRC 359
Db 281 -----KNTRTWFKL-----NDKLDYEC 297
QY 360 SDIF--BYRHSVCINGKNWPEVDCTEKREQCPPPPQIPNAQNMTTIVNYQDGKVAVLC 417
Db 298 LVGFENEYKHT----- 308
QY 418 KENYLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTSFPLSVPPGSGTYYR 477
Db 309 -----KGSITCTYVGSWSDTSCYDSTRTGCGPPPIDNGDITSLSLPEYELSSVDYQ 360
QY 478 COSFYKLOGSVTVTCRNKQWSEPPRCCLDPCYVSEENKNNIOLKWRNDGKLYAKTGDV 537
Db 361 CQYIILKGNKTITCRNGKWSSEPTCLHACVIPENIMEAHNIILKWRHTEKIYHSGEDI 420
QY 538 EFOCKFPHKAMISSPPFRAICQEGKFEYPTC 568
Db 421 EFECKRGYKARGSLPFRKTCINGTINYPTC 451

RESULT 10
Q91275
ID Q91275 PRELIMINARY; PRT; 1053 AA.
AC Q91275;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Complement regulatory plasma protein.
OS Paratubrax nebulifer (barred sand bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Serranidae; Paralabrax.
OX NCBI_TaxID=30873;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=94318039; PubMed=8042982;
RA Dahmen A., Kaidoh T., Zipfel P.F., Giggli I.;
RT "Cloning and characterization of a cDNA representing a putative
RT complement-regulatory plasma protein from barred sand bass (Paralabrax
RT nebulifer)."
RL Biochem. J. 301:391-397(1994).
DR EMBL; L21703; AAA92556.1;
DR HSSP; P08603; 1HFH.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 16.
DR SMART; SM00032; CCP; 16.
SQ SEQUENCE 1053 AA; 117597 MW; F27E32C3AD76D503 CRC64;

Query Match 15.0%; Score 476; DB 13; Length 1053;
Best Local Similarity 24.8%; Pred. No. 1.7e-33;
Matches 159; Conservative 94; Mismatches 237; Indels 152; Gaps 34;

QY 37 EDYNPFQVPT--TCEVFYISCEYNFVSPFTRITCTEGWSPPTKCL--RWC-SFFPV 92
Db 455 KDPNGFFRGYTGRLVLYTCQKYLFTGEGWAAKCVGVWPELTTCISNTTCGKFPFI 514
QY 93 KNGHSESSGLI--HLEGDTVQIICNTGYSLQNNKINISCVRGW-----STPPICSTKG 145
Db 515 PNAE-----VIRYPEVQTVVVICNQGYSTQAN--SFSCEGNWLLYGLSPQICTLRAD 567
QY 146 ECHVPPILEANVDAOPK-KESYKVGDLVKFSCRMLIRV-GSDSVQCYOGWSPNF----- 198
Db 568 VCGPPPEAENAVVTKSYQREYLSGSEVYLCRDKYIPLEGVDTITCRNGQWQKEIKCTSS 627
QY 199 -----PTCKQVRSRGPP 211
Db 628 CDKLDVMTDMFTADKEIYEGOTIRYQCLLDAGEGTATCNTKWKSPCK--VRFCELP 685
QY 212 POLSNGEYKEIRKEEYGHNEVVEYDNCNPNF-IINGPKKIQCVDGWT--LPTCVBOVKT 269

```



DR PRINTS: PR00453; VWFADOMAIN.  
 DR PRODOM; PD002153; Pentaxin; 1.  
 DR SMART; SM00032; CCP; 34.  
 DR SMART; SM00181; EGF; 15.  
 DR SMART; SM00179; EGF CA; 9.  
 DR SMART; SM00001; EGF-like; 3.  
 DR SMART; SM00159; PTH; 1.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 5.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_9.  
 DR PROSITE; PS01186; EGF\_2; 11.  
 DR PROSITE; PS01187; EGF CA; 6.  
 DR PROSITE; PS0234; VWFA; 1.  
 KW SIGNAL.  
 FT SIGNAL.  
 SQ SEQUENCE 3567 AA; 387391 MW; 8FBA8276E12293E5 CRC64;  
 Query Match 12.4%; Score 393.5; DB 11; Length 3567;  
 Best Local Similarity 24.8%; Pred. No. 1.6e-25;  
 Matches 152; Conservative 80; Mismatches 248; Indels 133; Gaps 39;  
 QY 26 PKIHGFLYDEEDYNPFSQVPTGEVYSC--YNFVSPSKSFWRITTCDEG-WSP-TP 81  
 Db 1693 PALENG-YSADFH-----AGSTVYQCTSGYLLGDSRMF-----CTDNGSWNGISP 1740  
 QY 82 KCLRM-----CSFFV-----KNGHSE 98  
 Db 1741 SCLDVBECAGSDCSEHASCLNTNGSVVSCNPPYTGDKRCAEPVKRAPENPENGH-- 1798  
 QY 99 SSGLIHLEGDTVOICNTGYSLQNNKNTSCVBERG-WS-TTPICSTFKGCHVPILAN- 155  
 Db 1799 SSGEITVGAVTFSCDEGHELV-GVSTITCLETGEMDLRPSCEAI--SCGVPPYPENG 1855  
 QY 156 -VDAQPKKESYKVGDLKFSCKRNLIRVGSDSYQYQFG-WSPNFTCTCKGVRSCGPPQ 213  
 Db 1856 GVDG-----SAFTYGVVYRCDKGYTLSGDEESACLASGWSHSPVC--ELVKCSQPED 1909  
 QY 214 LSNGEVKEIKERY-----GHNEVVEYDCNPNIINGPKKIQC-VDGEWTL-PTCQEYV 266  
 Db 1910 INNG-----KYILSGLTYLSIASYSCENGYSLOGPSLLECTASGSDRAPPSG--QL 1959  
 QY 267 KTCGYIPELEYGVQVSPYPYQHGVSVEYCNREYAMIGNMITC-INGIW-TELPWCVA 324  
 Db 1960 VSCGEPPIKDAVITGS--NFTGTNTVAYTCREGYTLAGPDTIVCOANGKNSNHOCLA 2017  
 QY 325 THQKRCKIAGVNIKTLKLSGKEFNHNSRIRYRCSDFIRYHS---VC-INGKWNPEVD 380  
 Db 2018 VSCDEPPNVDHASPETAHRLFG-----DTAFYICADGYSLADNSOLICNAQGNWVPAG 2071  
 QY 381 CTEKR--EQFCPPPPPOIP-NAQNMNTTVNYQGEKVAVLCKENYLLPEAKEIVC-KDGRW 436  
 Db 2072 QAVPRCIAHFCEKPPSVSYSILESVSFAKFAAGSVVSKMEGVFLNTSAKIECLRGGEW 2131  
 QY 437 QSLPRCVES-TAVCGPPPPINNGDTTSEPLSV-YPPGSTVTVYRCQSFYKLOGSVTVTCR- 493  
 Db 2132 SPSPLSVQICPVKCGPPPSIANG-----YPSGTYNSGAVVAVSCHGPKYIKGKSTCEA 2187  
 QY 494 NKOWSPPRCLDPCVYSEENMKNNIQLKWRNDGKLYAKTGDAVE----FQCKFPKAMI 549  
 Db 2188 TGOWSKPTTCHPVCNPPKVEN-----GFLHTTGTTFEFAFQCPNGYKA-A 2237  
 QY 550 SSPPPFAICQCK 562  
 Db 2238 GSPVF--VCOANR 2248  
 RESULT 13  
 Q61405 PRELIMINARY; PRT; 303 AA.  
 AC Q61405;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Complement factor H-related protein.  
 GN A1194596.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=90153969; PubMed=1689298;  
 RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,  
 RA Chaplin D.D.;  
 RT Identification and sequence analysis of 4 complement factor H-related  
 RT transcripts in mouse liver.\*;  
 RL J. Biol. Chem. 265:3193-3201(1990).  
 DR EMBL; M29007; AAA37413.1; -.  
 DR HSSP; P10998; LVVD  
 DR MGD; MGI:2138169; A1194596  
 DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam; PF00084; sushi; 4.  
 DR SMART; SM00032; CCP; 4.  
 SQ SEQUENCE 303 AA; 34498 MW; 51C66E50906F4C24 CRC64;  
 Query Match 12.2%; Score 387.5; DB 11; Length 303;  
 Best Local Similarity 25.9%; Pred. No. 2.7e-26;  
 Matches 83; Conservative 38; Mismatches 83; Indels 117; Gaps 4;  
 QY 1 MLLFSVILISWSTVYGEG----- 20  
 Db 6 MLLSNILLTAWLSTAKEGKCTSPPYILNGIYTPHRIIHKSDDEIRYECNYGFYVTGS 65  
 QY 21 -----TL--CDPPKIHGHGLYDEEDYNPFSQVPTGEVYSCYENFVSP 62  
 Db 66 TVSKCTPTGWIPVPRCTLAPCEFPQFKYGRLYEESLRPNFPVSGNKYSYKCDNGFSPP 125  
 QY 63 SKSFWRITCTEGSGWSPKCLRMCSFPFKNGHSESSGLIHLEGTVOICNTGYSLON 122  
 Db 126 SGYSWDYLRCTAQGWEPVCKRVKVFHYVENDSTYWEKIYVQGSGLKVCQCYNGYSLO 185  
 QY 123 NEKNISCVRGWSTPTPICSTFKGCHVPILANVDAQPKKESYKVGDLKFSCKRNLIRV 182  
 Db 186 GQDTMTCTENGWSPPKC----- 206  
 QY 183 GDSVQCYQFGWSPNPTCKGVRSCGPPQPSLNGEVEKEIRKEEYGHNEVYDCNPFI 242  
 Db 207 NS-----TRTCGPPPPIDNGDITSLSLPEYELPSVDYQCKYKL 246  
 QY 243 INGPKKIQCVGEGWTLTPCV 263  
 Db 247 LKGNKTIITCRNGWSEPPMCL 267  
 RESULT 14  
 Q29528 PRELIMINARY; PRT; 1911 AA.  
 AC Q29528;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Complement component receptor type 1 (Fragment).  
 GN CRI.  
 OS Papio hamadryas (Hamadryas baboon).  
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;  
 OC Cercopitheciinae; Papio.  
 OX NCBI\_TaxID=9557;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Clemenza L., Subramanian B.V., Nickells M.W., Hourcade D.E.,  
 RA Atkinson J.P.;  
 RT "Primary sequence of the baboon 200 kDa C3b/C4b receptor (CRI).";  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; L39791; AAA62170.1; -.  
 DR HSSP; P08603; LHFI.



Search completed: July 17, 2003, 13:53:00  
Job time : 91 secs

---

GenCore version 5.1.6

Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 13:49:33 ; Search time 24 Seconds  
(without alignments)  
983.335 Million cell updates/sec

Title: US-09-805-337A-2

Perfect score: 3183

Sequence: 1 MLLFSVLISWSTVGEG.....SSPPFRAICGKFEYPICE 569

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1538	48.3	1231	1 CFAH_HUMAN	P08603 homo sapien
2	1427.5	44.8	1234	1 CFAH_MOUSE	P06909 mus musculus
3	1050.5	33.0	330	1 FHR1_HUMAN	Q03591 homo sapien
4	927.5	29.1	270	1 FHR2_HUMAN	P36980 homo sapien
5	861	27.0	685	1 CFAH_BOVIN	Q28085 bos taurus
6	783	24.6	668	1 F13B_MOUSE	Q07968 mus musculus
7	764.5	24.0	661	1 F13B_HUMAN	P05160 homo sapien
8	723.5	22.7	330	1 FHR3_HUMAN	Q02985 homo sapien
9	681	21.4	331	1 FHR4_HUMAN	Q92496 homo sapien
10	411	12.9	1033	1 CR1_HUMAN	P20023 homo sapien
11	385	12.1	2039	1 CR2_HUMAN	P17927 homo sapien
12	382	12.0	1025	1 CR2_MOUSE	P19070 mus musculus
13	339.5	10.7	558	1 C4BP_RAT	Q63514 rattus norv
14	328	10.3	610	1 C4BP_BOVIN	Q28065 bos taurus
15	303	9.5	597	1 C4BP_HUMAN	P04003 homo sapien
16	300.5	9.4	469	1 C4BP_MOUSE	P08607 mus musculus
17	281	8.8	830	1 LEM3_HUMAN	P16109 homo sapien
18	274.5	8.6	612	1 LEM2_MOUSE	Q00690 mus musculus
19	264.5	8.3	610	1 LEM2_HUMAN	P16581 homo sapien
20	255	8.0	768	1 LEM3_MOUSE	Q01102 mus musculus
21	253	7.9	768	1 LEM3_RAT	P98106 mus musculus
22	252.5	7.9	769	1 LEM3_SHEEP	P98109 ovis aries
23	247.5	7.8	646	1 LEM3_BOVIN	P42201 bos taurus
24	243.5	7.7	551	1 LEM2_RABIT	P27113 oryctolagus
25	243	7.6	345	1 APOH_MOUSE	Q01339 mus musculus
26	237.5	7.5	345	1 APOH_CANFA	P33703 canis fami
27	235.5	7.4	611	1 LEM2_CANFA	P33730 canis fami
28	233	7.3	549	1 LEM2_RAT	P98105 rattus norv
29	231.5	7.3	263	1 VCP_VACCV	P10998 vaccinia vi
30	231	7.3	345	1 APOH_HUMAN	P02749 homo sapien
31	225.5	7.1	360	1 CCPH_HSVSA	Q01016 herpesvirus
32	224.5	7.1	345	1 APOH_BOVIN	P17690 bos taurus
33	224	7.0	507	1 DAF_CAVPO	Q60401 cavia porce

## RESULT 1

ID	CFAH_HUMAN	STANDARD;	PROT;	1231 AA.
AC	P08603: Q14570; P78435; Q9NU86;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Complement factor H precursor (H factor I).			
GN	HF1 OR HF OR CFH.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT Y-402.			
RC	TISSUE=Liver;			
RX	MEDLINE=88134059; PubMed=2963625;			
RA	Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;			
RT	"The complete amino acid sequence of human complement factor H.;"			
RL	Biochem. J. 249:593-602(1988).			
RN	[2]			
RP	SEQUENCE OF 53-445 FROM N.A.			
RX	MEDLINE=87054207; PubMed=2946589;			
RA	Schulz T.F., Schwaeble W., Stanley K.K., Weiss E., Dierich M.P.;			
RT	"Human complement factor H: Isolation of cDNA clones and partial cDNA sequence of the 38-kDa tryptic fragment containing the binding site for C3b.;"			
RL	Eur. J. Immunol. 16:1351-1355(1986).			
RN	[3]			
RP	SEQUENCE OF 226-449 FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE=86169701; PubMed=2937845;			
RA	Kristensen T., Wetsel R.A., Tack B.F.;			
RT	"Structural analysis of human complement protein H: homology with C4b binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.;"			
RL	J. Immunol. 136:3407-3411(1986).			
RN	[4]			
RP	SEQUENCE FROM N.A. (ISOFORM 2).			
RX	Submitted (JAN-2000) to the EMBL/GenBank/DBDJ databases.			
RN	[5]			
RP	SEQUENCE OF 1047-1231 FROM N.A.			
RX	MEDLINE=91201892; PubMed=1826708;			
RA	Estaller C., Koistinen V., Schwaeble W., Dierich M.P., Weiss E.H.;			
RT	"Cloning of the 1.4-kb mRNA species of human complement factor H reveals a novel member of the short consensus repeat family related to the carboxy terminal of the classical 150-kDa molecule.;"			
RL	J. Immunol. 146:3190-3196(1991).			
RN	[6]			
RP	SEQUENCE OF 19-35.			
RX	MEDLINE=83048213; PubMed=6215918;			
RA	Sim R.B., Discipio R.G.;			
RT	"Purification and structural studies on the complement-system control protein beta 1H (Factor H).;"			
RL	Biochem. J. 205:285-293(1982).			
RN	[7]			
RP	SEQUENCE OF 1-19 FROM N.A.			

P28175 tachypleus  
Q26422 carcinoscor  
Q8byh1 homo sapien  
P26644 rattus norv  
Q09101 drosophila  
Q13219 homo sapien  
P21115 vaccinia vi  
P08174 homo sapien  
Q61475 mus musculus  
Q01227 vaccinia vi  
P24083 vaccinia vi  
P49457 pongo pygma





P06909;  
01-JAN-1988 (Rel. 06, Created)  
01-JAN-1988 (Rel. 06, Last sequence update)  
15-JUN-2002 (Rel. 41, Last annotation update)  
Complement factor H precursor (protein beta-1-H).  
HFI OR CPH.  
Mus musculus (Mouse).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE-86233353; PubMed-2940596;  
Kristensen T.; Tack B.F.;  
"Murine protein H is comprised of 20 repeating units, 61 amino acids  
in length.";  
Proc. Natl. Acad. Sci. U.S.A. 83:3963-3967(1986).  
[2]  
SEQUENCE OF 1-19 FROM N.A.  
STRAIN-BALB/c;  
MEDLINE-90148935; PubMed-253512;  
Munoz-Canoves P.; Tack B.F.; Vik D.P.;  
"Analysis of complement factor H mRNA expression: dexamethasone and  
IFN-gamma increase the level of H in L cells.";  
Biochemistry 28:9891-9897(1989).  
[3]  
SEQUENCE OF 1-18 FROM N.A.  
MEDLINE-90111033; PubMed-2136885;  
Natsume-Sakai S.; Nonaka M.; Nonaka M.; Harada Y.N.; Shreffler D.C.;  
Moriwaki K.;  
"Demonstration of an unusual allelic variation of mouse factor H by  
the complete cDNA sequence of the H.2 allotype.";  
J. Immunol. 144:358-362(1990).  
-1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF  
C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE  
C3BB COMPLEX (C3 CONVERTASE) AND THE (C3b)NBB COMPLEX (C5  
CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.  
-1- POLYMORPHISM: TWO CODOMINANT ALLELES OF FACTOR H ARE PRESENT IN  
MICE.  
-1- SIMILARITY: CONTAINS 20 SUSHI (SCR) DOMAINS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
-----  
EMBL; M12660; AAA37759.1; -;  
EMBL; J02891; AAA37795.1; -;  
EMBL; M31979; AAA37762.1; -;  
PIR; A26154; NEMSH.  
HSP; P08603; LHFI.  
MGD; MGI:88385; Cfh.  
InterPro; IPR000436; Sushi\_SCR\_CCP.  
Pfam; PF00084; sushi; 20.  
SMART; SM00032; CCP; 20.  
Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;  
Signal.  
-----  
1 18  
BY SIMILARITY.  
FT FT CHAIN 19 1234  
FT FT COMPLEMENT FACTOR H.  
FT FT 20 81  
FT FT DOMAIN 20 81  
FT FT 84 142  
FT FT DOMAIN 145 206  
FT FT DOMAIN 209 263  
FT FT DOMAIN 266 321  
FT FT DOMAIN 324 386  
FT FT DOMAIN 388 443  
FT FT DOMAIN 447 506  
FT FT DOMAIN 508 565  
FT FT DOMAIN 568 623  
FT FT DOMAIN 628 684  
FT FT DOMAIN 628 684

Qy	141	-----SFTKG-----	145
Db	505	CIKSCDMPVENSITKNTWFKLNDKLDVECLVGENEYKHTGTSITYYGNSDTPSC	564
Qy	146	---ECHYPILEANVDAOPKESYKVGDLAFSCRKNLIRVSGDSVQCYQFGWSPNPTCK	202
Db	565	YERCSVPTLDRKLWSPREKRYRGVGLLFECSHG-HRVGPDVQCYHFGWSPGPTCK	623
Qy	203	GOVRSCGPPQLSNGEYKETRKEEYGHNEVVEYDCNPFLINGPKKIQCVGDEWTTLPCT	262
Db	624	GOVASCAPPEILNGBEINGAKKVEYSHGEVVKYDCKPFTLLKGPNKIQCYDGNWTTLPVC	683
Qy	263	VEQVKTCGYIPELEYGVQVSPVPPYOHGVSVNECRNEYAMIGNNITCINGIHWTELPMC	322
Db	684	IEEBTCDGDIPELEHSGAKSVPPYHHGDSVEFTICEENFTMIGHGVSISGKWTQLPKC	743
Qy	323	VATHQLKRCKIAGVNTIKLLKLGKGFENHNSRIRYRCSDIFRYRHVSCINGKKNPVDCT	382
Db	744	VATDLEKCRVLKSTGTEAKPKLITFTHNSTWDYKCRQKQEYERSICINGKWDPEPNC	803
Qy	383	EKRBQFCPPPPQIPNQNAMTNTVYQDGEKAVVLCKENYLLPBAKEIVCKDGRWQSLPRC	442
Db	804	SKTS--CPPPPQIPNTQVIETTVKYLDEKLSVLCQDNYLTQDSEENVCKDGRWQSLPRC	861
Qy	443	VE-----	444
Db	862	IEIKPCSQPTTIEHGSINLPSREERRDSIESSSHEHGTTFSVYCDGDFRIPEENRITCY	921
Qy	445	-----	444
Db	922	MGRWSTPPRCVGLPCGPPPSPIPLGTVLSLELESYQHGEVYTHCSTGFGIDGPAFICEGG	981
Qy	445	-----	444
Db	982	KWSDPPKCIKTDQDLVLTVKNAIIRGSKSKSYRTGEQVTFRCQSPYQMNGSDTVCVNSR	1041
Qy	445	-----	444
Db	1042	WIGQPVCKDNCSDVDPHPVNATVTTKNKYLHGDRVRYEKNKPLELFCQVEVMECNGTW	1101
Qy	445	-----STAYCGPPPSINNGDTSFPLSVYPPGSTVTVYRCQSPYKLGQSVTVTCRNKQ	496
Db	1102	TEPKCRDSTGKCGPPPDINGDITSLSLPVYELPSVEYQCYKYLLGKKTITCTNCK	1161
Qy	497	WSEPPRCLDPCVSEBEMKNNTQLKWRNDGLKYAKTGDAVFQCKFPPIKAMLSPPFPRA	556
Db	1162	WSEPPRTLACVPIPENIMESHNTILKWRHTEIYSHSGEDIEFGCKYGYKYKARDSPPERT	1221
Qy	557	ICQEGKFEYPIG	568
Db	1222	KCINGTINPTC	1233
RESULT 3			
FHR1_HUMAN			
ID	FHR1_HUMAN	STANDARD;	PRT; 330 AA.
AC	Q03591; Q90U17;		
DT	01-JUN-1994 (Rel. 29, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Complement factor H-related protein 1 precursor (FHR-1) (H factor-like		
DE	protein 1) (H-factor like 1) (H36).		
GN	HFL1 OR FHR1 OR CFHL.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.92.		
RC	TISSUE=Liver;		
RX	MEDLINE=91201892; PubMed=1826708;		
RA	Estaller C., Koistinen V., Schwaebel W., Dierich M.P., Weiss E.H.;		



01-JUN-1994 (Rel. 29, Last sequence update)  
 15-JUN-2002 (Rel. 41, Last annotation update)  
 Complement factor H-related protein 2 precursor (FHR-2) (H factor-like protein 2) (H factor-like 3) (DDESK59).  
 HFL3 OR FHR2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 [1]  
 SEQUENCE FROM N.A.  
 RQ TISSUE=Liver;  
 RC MEDLINE=92251200; PubMed=1533657;  
 RX Skerka C., Hilmann C., Horstmann R.D., Zipfel P.E.;  
 RA "Two additional human serum proteins structurally related to complement factor H. Evidence for a family of factor H-related genes.";  
 RT J. Immunol. 148:3313-3318(1992).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=94226679; PubMed=8172644;  
 RX Zipfel P.F., Skerka C.;  
 RA "Complement factor H and related proteins: an expanding family of complement-regulatory proteins";  
 RT Immunol. Today 15:121-126(1994).  
 RL [1]  
 CC -/- FUNCTION: MIGHT BE INVOLVED IN COMPLEMENT REGULATION. CAN ASSOCIATE WITH LIPOPROTEINS AND MAY PLAY A ROLE IN LIPID METABOLISM.  
 CC [1]  
 CC -/- SUBCELLULAR LOCATION: Extracellular.  
 CC [1]  
 CC -/- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A TRUNCATED FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.  
 CC [1]  
 CC -/- PTM: N-GLYCOSYLATED. TWO FORMS ARE OBSERVED; ONE WITH A SINGLE SIDE CHAIN, THE OTHER IS NOT GLYCOSYLATED.  
 CC [1]  
 CC -/- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.  
 CC [1]  
 CC -/- SIMILARITY: STRONG, TO FACTOR H.  
 CC [1]  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC [1]  
 DR EMBL; X64877; CAA6096.1; .  
 DR EMBL; X86564; CAA60375.1; .  
 DR EMBL; X86565; CAA60375.1; JOINED.  
 DR EMBL; X86566; CAA60375.1; JOINED.  
 DR EMBL; X86567; CAA60375.1; JOINED.  
 DR PIR; S24564; S24564.  
 DR KSSP; P08603; IHFI.  
 DR Genew; HGNC:4890; HFL3.  
 DR MIM; 600889; .  
 DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam; PF00084; sushi; 4.  
 DR SMART; SM00032; CCP; 4.  
 KW Repeat; Glycoprotein; Sushi; Signal; Alternative splicing.  
 FT SIGNAL 1 18  
 FT CHAIN 19 270  
 FT DOMAIN 22 84  
 FT DOMAIN 86 141  
 FT DOMAIN 148 204  
 FT DOMAIN 206 268  
 FT DISULFID 23 72  
 FT DISULFID 55 83  
 FT DISULFID 87 129  
 BY SIMILARITY.  
 BY SIMILARITY.

FT DISULFID 114 140 BY SIMILARITY.  
 FT DISULFID 149 192 BY SIMILARITY.  
 FT DISULFID 178 203 BY SIMILARITY.  
 FT DISULFID 207 257 BY SIMILARITY.  
 FT DISULFID 241 267 BY SIMILARITY.  
 FT CARBOHYD 126 126 N-LINKED (GLCNAC...).  
 FT VARSPLIC 144 171 ISAEKCGPPPIDNGDITSLLSYAPG -> S (IN TRUNCATED ISOFORM).  
 FT CONFLICT 85 85 R -> K (IN REF. 2).  
 SQ SEQUENCE 270 AA; 30651 MW; ELB2E9F139B217A8 CRC64;  
 Query Match 29.1%; Score 927.5; DB 1; Length 270;  
 Best Local Similarity 34.6%; Pred. No. 3.1e-61;  
 Matches 198; Conservative 20; Mismatches 47; Indels 307; Gaps 3;  
 QY 1 MLLFSLVILISWVTVGGECTLDPFKIHGHFLYDEEDYNPFQSVPTGVEFYVSCENFY 60  
 DB 1 MWLLSVILISRISSVGGGEAMFCDFPKINHGIYDEKYPKFFSQVPTGVEFYVSCENFY 60  
 QY 61 SPSKSEFWTRITCTEEGWSPTPKLRMCSPFVKNGHSESSGLIHLEGDVQIICNTGYSL 120  
 DB 61 SPSKSEFWTRITCAEKGWSPTPKLRCLCFPFVENGHSESSQTHLEGDVQIICNTGYRL 120  
 QY 121 QNNEKNISCVERGWSPTPKISFTKGECHVPILANVDAQPKKESYKVGVDVLKFSCKRNLI 180  
 DB 121 QNNEKNISCVERGWSPTPKISFTKGECHVPILANVDAQPKKESYKVGVDVLKFSCKRNLI 180  
 QY 181 RVGSDSVQCVQFGWSPNFTCKGOVRSCTGPPQLSNCEYKEIRKEEYGHNEVVEYDCNPN 240  
 DB 141  
 QY 241 FIINGPKKIQCVGDGWTTLPTCQEYQVKTGCIPELEYGVQPSVPPYQHGVSVEVNCNE 300  
 DB 141  
 QY 301 YAMIGNMTCINGIWTCLPMCVATHQLARCKIAGVNIKTLKLSKEFNHNSRIYRCS 360  
 DB 141  
 QY 361 DIPRYRHSVINGKNWPEVDCTEKREQFCPPPPQIPNAQNMNTTVNYQGEKVAVLCKEN 420  
 DB 141  
 QY 421 YLLPEAKETVCKDGRWQSLPRCVSTAYCGPPPSINNGDTSPLSVYPPGSTVTVYRCS 480  
 DB 141  
 QY 481 FYKLGQSVTVTCENKQWSEPPRCCLDPVCVYSEENKANNIOLKWRNDGKLYAKTGDAVEFQ 540  
 DB 181 LYQLEGNNQITCRNGQWSEPPRCCLDPVCVYSEENKANNIOLKWRNDGKLYAKTGDAVEFQ 540  
 QY 541 CK---FPHKAMISSPPFRAICQEGKPEYPICE 569  
 DB 241 CASGYHPTK-----SHSFRAMCQNGKLVYPSCE 268  
 RESULT 5  
 CFAH\_BOVIN STANDARD; PRT; 685 AA.  
 ID CFAH\_BOVIN  
 AC Q28085;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Complement factor H (H factor 1) (Fragments).  
 GN HFI1  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE OF 17-685 FROM N.A., AND SEQUENCE OF 1-16.  
 RC TISSUE=Liver;



KW Plasma: Blood coagulation; Repeat; Glycoprotein; Signal; Sushi.  
 FT SIGNAL 1 20 BY SIMILARITY.  
 FT CHAIN 21 668 COAGULATION FACTOR XIII B CHAIN.  
 FT DOMAIN 24 88 SUSHI 1.  
 FT DOMAIN 90 147 SUSHI 2.  
 FT DOMAIN 152 209 SUSHI 3.  
 FT DOMAIN 212 268 SUSHI 4.  
 FT DOMAIN 273 328 SUSHI 5.  
 FT DOMAIN 335 390 SUSHI 6.  
 FT DOMAIN 395 451 SUSHI 7.  
 FT DOMAIN 453 516 SUSHI 8.  
 FT DOMAIN 523 579 SUSHI 9.  
 FT DOMAIN 581 647 SUSHI 10.  
 FT DISULFID 25 76 BY SIMILARITY.  
 FT DISULFID 59 87 BY SIMILARITY.  
 FT DISULFID 91 135 BY SIMILARITY.  
 FT DISULFID 118 146 BY SIMILARITY.  
 FT DISULFID 153 197 BY SIMILARITY.  
 FT DISULFID 180 208 BY SIMILARITY.  
 FT DISULFID 213 255 BY SIMILARITY.  
 FT DISULFID 241 267 BY SIMILARITY.  
 FT DISULFID 274 316 BY SIMILARITY.  
 FT DISULFID 302 327 BY SIMILARITY.  
 FT DISULFID 336 378 BY SIMILARITY.  
 FT DISULFID 364 389 BY SIMILARITY.  
 FT DISULFID 396 439 BY SIMILARITY.  
 FT DISULFID 425 450 BY SIMILARITY.  
 FT DISULFID 454 505 BY SIMILARITY.  
 FT DISULFID 486 515 BY SIMILARITY.  
 FT DISULFID 524 567 BY SIMILARITY.  
 FT DISULFID 553 578 BY SIMILARITY.  
 FT DISULFID 582 636 BY SIMILARITY.  
 FT DISULFID 616 646 BY SIMILARITY.  
 FT CARBOHYD 162 162 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 545 545 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 668 AA; 80BC9E00A9E53FA6 CRC64;

Query Match 24.6%; Score 783; DB 1; Length 668;  
 Best Local Similarity 28.7%; Pred. No. 3.6e-50;  
 Matches 187; Conservative 95; Mismatches 266; Indels 104; Gaps 18;

QY 7 VILISVWTVGGRTGLCDPKIHG---FLVD-EEDYDPFSQVPTGEVFFVYSCENFVS 61  
 DB 9 ILLLLSGELYAEKQCDPTVNGRIAQYVYTFKSFYFPMs---VDDKLSFFCLAGIAT 65  
 QY 62 PSKSFWRITCTEGNSPTPKLRLMCSFPVKNHSESSGLIHLEGTQVLIICNTGYSLQ 121  
 DB 66 ESKGQEQIRCTAEGNSPNPRCKKLPDLRNGVYVNDKLVYKLPQRMNSYGCSSGYKIT 125  
 QY 122 --NNEKNISCVRGNSPTPSCSTKGECHVPILP-----ANVDAQPKESYKVGDLKFSCKKNL 179  
 DB 126 GGDKEEVVHCLSGAGSSQPSCKRKEQETCLAPELEHGNYSSTORTFFKVDIVAYTCAGYY 185  
 QY 154 -----ANVDAQPKESYKVGDLKFSCKKNL 179  
 DB 186 TTTGKTGTGAECQANGSWLTPQCNKLMCSRLRIENGYFHPVQTYEEDGLVOFFCHENY 245  
 QY 180 IRVGSQVQYQFGSNFPPTCKGQVRSGPPQLSNGEVKTEKREYGHNEVEYDCNP 239  
 DB 246 YLSGSLIQYNGFYWPSPICEGRNRNCPVPVPLNSKIOP-HSTYRHRGERVHLECEL 304  
 QY 240 NFIINGPKIQVDGWTLPICVEQVK--TCGYIPELYGVQPSVPYQHGVSVEVNC 297  
 DB 305 NFVIOGSELLCNGKWTPEPCIEBEKVACEQPPSVENGVAHPHSEIYYSQDKVYRC 364  
 QY 298 RNEYAMIGNMTTCINGIWTCLPMCVATHOLARKC---IA-GVNKTLKLSGKGFNNH 352  
 DB 365 GGYSLRGSTTCNRRGRTLPPECV--ENIENCKPPDPIANGVVVDGLL-----ASYTGG 418  
 QY 353 SRIYRCSDIFRYHSV---CINGKNWPEVDCTEKREQCPPPQTPNAQNMNTTVNYQ- 408  
 DB 419 SSVEYRCNEYLLKGSSETRCBEGAWSSPPVCLP-----CTIDVDMNRNRLQKWKYEG 474

QY 409 ---DGEKVAVLCKENYLLPEA---KEIVCKDGRWQ-SLPRCV--ESTAYGPPPSINNGD 459  
 DB 475 KILHGLDIDFVCKQGYNLSPISLSEISAQCNRGDVRYPMCIRKESKGMCAASPPVIRNGD 534  
 QY 460 TTFEPLSVYPPGSTVTVYRCOSFYKLGQSVTVTCRNKOWSEPPRCRLDPCVYVSEENNNKNI 519  
 DB 535 IVSSAARTYENGSSVEYRCFNDHFLQGSQNVYCDGVMTTPPSCLEPCTLSFVEMDKNYL 594  
 QY 520 QLKWRNDGKLYAKTGDVAVEFOCK---PPHKAMISSPPFRAICOGREYEPIC 568  
 DB 595 QLKWNFDNRPLILHGEVIEFCKRDYISETSAGSVLRVQCDRGLKYPKC 646

## RESULT 7

FL3B\_HUMAN  
 ID FL3B\_HUMAN STANDARD; PRT; 661 AA.  
 AC P05160;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Coagulation factor XIII B chain precursor (Protein-glutamine gamma-glutamyltransferase B chain) (Transglutaminase B chain) (Fibrin stabilizing factor B subunit).  
 DE FL3B.  
 GN Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91105054; PubMed-2271707;  
 RA Botteneus R.E., Ichinose A., Davie E.W.;  
 RT "Nucleotide sequence of the gene for the b subunit of human factor XIII";  
 RL Biochemistry 29:11195-11209(1990).  
 RN [2]  
 RP SEQUENCE OF 2-661 FROM N.A.  
 RX MEDLINE-87026535; PubMed-3021194;  
 RA Ichinose A., McMullen B.A., Fujikawa K., Davie E.W.;  
 RT "Amino acid sequence of the b subunit of human factor XIII, a protein composed of ten repetitive segments";  
 RL Biochemistry 25:4633-4638(1986).  
 RN [3]  
 RP REVISIONS.  
 RA Ichinose A.;  
 RL Submitted (FEB-1987) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 1-20 FROM N.A.  
 RX TISSUE=Liver;  
 RX MEDLINE-90251467; PubMed-2339067;  
 RA Grundmann U., Nerlich C., Rein T., Zettlmeissl G.;  
 RT "Complete cDNA sequence encoding the B subunit of human factor XIII";  
 RL Nucleic Acids Res. 18:2817-2817(1990).  
 RN [5]  
 RP VARIANT PHE-450.  
 RX MEDLINE-93313189; PubMed-8324218;  
 RA Hashiguchi T., Saito M., Morishita E., Matsuda T., Ichinose A.;  
 RT "Two genetic defects in a patient with complete deficiency of the b-subunit for coagulation factor XIII";  
 RL Blood 82:145-150(1993).  
 CC -!- FUNCTION: THE B CHAIN OF FACTOR XIII IS NOT CATALYTICALLY ACTIVE, BUT IS THOUGHT TO STABILIZE THE A SUBUNITS AND REGULATE THE RATE OF TRANSGLUTAMINASE FORMATION BY THROMBIN.  
 CC -!- SUBUNIT: TETRAMER OF TWO A CHAINS AND TWO B CHAINS.  
 CC -!- DISEASE: Defects in FL3B can result in a lifelong bleeding tendency, defective wound healing, and habitual abortion.  
 CC -!- SIMILARITY: CONTAINS 10 SUSHI (SCR) DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial



"Complement factor H and related proteins: an expanding family of complement-regulatory proteins"; Immunol. Today 15:121-126(1994).

-|- FUNCTION: MIGHT BE INVOLVED IN COMPLEMENT REGULATION.

CC -|- SUBCELLULAR LOCATION: Extracellular.

CC -|- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.

CC -|- SIMILARITY: STRONG, TO FACTOR H.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; X68679; CAA48639.1; -  
DR EMBL; AL049741; CAB53064.1; -  
DR PIR; A45222; A45222.  
DR HSSP; P10998; 1VVD.  
DR MIM; 605336; -  
DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
DR Pfam; PF00084; sushi; 5.  
DR SMART; SM00032; CCP; 4.  
DR Repeat; Glycoprotein; Sushi; Signal.  
FT SIGNAL 1 18 POTENTIAL.  
FT CHAIN 19 330 COMPLEMENT FACTOR H-RELATED PROTEIN 3.  
FT DOMAIN 22 84 SUSHI 1.  
FT DOMAIN 86 141 SUSHI 2.  
FT DOMAIN 145 204 SUSHI 3.  
FT DOMAIN 209 265 SUSHI 4.  
FT DOMAIN 267 330 SUSHI 5.  
FT DISULFID 23 72 BY SIMILARITY.  
FT DISULFID 55 83 BY SIMILARITY.  
FT DISULFID 87 129 BY SIMILARITY.  
FT DISULFID 114 140 BY SIMILARITY.  
FT DISULFID 146 192 BY SIMILARITY.  
FT DISULFID 175 203 BY SIMILARITY.  
FT DISULFID 210 253 BY SIMILARITY.  
FT DISULFID 239 264 BY SIMILARITY.  
FT DISULFID 268 319 BY SIMILARITY.  
FT DISULFID 302 329 BY SIMILARITY.  
FT CARBOHYD 108 108 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 185 185 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 205 205 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 309 309 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 123 123 A -> VR (IN REF. 1).  
FT CONFLICT 132 132 K -> N (IN REF. 1).  
FT CONFLICT 143 143 Q -> D (IN REF. 1).  
FT CONFLICT 194 194 Q -> R (IN REF. 1).  
FT CONFLICT 241 241 P -> S (IN REF. 1).  
FT CONFLICT 260 260 E -> A (IN REF. 1).  
SQ SEQUENCE 330 AA; 37323 MW; 67B64121D71CD65F CRC64;

Query Match	22.7%	Score 723.5;	DB 1;	Length 330;
Best Local Similarity	26.9%	Pred. NO. 3.8e-46;		
Matches 153; Conservative	59;	Mismatches 118;	Indels 239;	Gaps 4;

  

QY	1	MLLFFSVILISWSTVGGECTL	CDPKPIHGGELVDEEDYNPF	SOVPTGEVFFYSCEYNFV	60
Db	1	MLLLINVILTWLWSCANGQV	KCDFFDIKHGGLFHENMR	PPYFPAVGKYSYICDEHFE	60
QY	61	SPSKSWTRITCTEEGWSPT	PCLRCMSPFFVKNHSGESGL	THLEGDTVQIICTGYSL	120
Db	61	TPSGSYWDYIHTQNGWSAP	VCPLRCRKYFPYLENGYNQY	GKRKFVQGNSTEVACHPGYGL	120
QY	121	QNEKNMISCEVRGWSPTT	ICSTFKGECHVPILLEANV	DAQPKKBSYKVGVDLFCRK	NLI 180
Db	121	PRAQTTVTCTEGWSP	TPRCI		141
QY	181	RVGSDSVQCYQFCWSPNF	FTCKGOVRSQPPQLSNGE	VKEIKRKEYGHNEVVEYDC	NP 240
Db	142	-----RVKTC-----	-----SKSDI-----		151

Qy	241	FIINGPKKIQCVGDGWTTLPTCVBOVKTCGYIPELEYGYVOPSPYQHGVSVEYNCRNE	300
Db	152	-----BIENGFISSSSIIYLNRKIOYKCRPG	178
Qy	301	YAMIGNNMITCINGIWTLPVCVATHQLKCKIAGVNIKTLKLSGKKEFNHSRIYRCS	360
Db	179	YATADGN-----	185
Qy	361	DIFRYRHISVCINGKWNPEVDCTEKREQCPPPOIPNAQNMTTIVNYDQGEKVAVLCKEN	420
Db	186	-----	185
Qy	421	YLLPEAKEIYCKDGRWQSLPRCVSEYATYCGPPPSINNGDTTSFPLSVYPGSTVYRCQS	480
Db	186	-----SSGSITCLONGWCSAQPICINSEKCGPPPIISNGDTSFLLKVYVPSRVEYQCQP	241
Qy	481	FYKLOGSVTVTCRNKONSEPPRCCLDPCVSENNKNINQLKRWNDCKLYAKTGDAVEFQ	540
Db	242	YYELOGSVTVTCNSGEWSEPPRCIHPCIIITEENKNKNIKLGRSDRKYAKTGDTIEFM	301
Qy	541	CKFPHKAMISSPPRAICQBGKFEYPICE	569
Db	302	CKLGYNANTSILSFQAVCRGIVYPRCE	330

## RESULT 9

```

RESULT 9
FHR4_HUMAN
ID FHR4_HUMAN STANDARD; PRT; 331 AA.
AC Q92496; Q9UJY6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement factor H-related protein 4 precursor (FHR-4).
GN FHR4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE-97190290; PubMed-9038172;
RX Skerka C., Hellwege J., Weber W., Tilkorn A., Buck F., Marti T.,
RA Kampen E., Beisiegel U., Zipfel P.F.;
RT "The human factor H-related protein 4 (FHR-4). A novel short consensus
RT repeat-containing protein is associated with human triglycylide-rich
RT lipoproteins."
RL J. Biol. Chem. 272:5627-5634(1997).
RN [2]
RP SEQUENCE OF 1-19 FROM N.A.
RA Male D.A., Ormsby R.J., Giannakis E., Gordon D.L.;
RT "Promoter region of complement factor H-related 4 (FHR-4) gene."
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RX MEDLINE-98136603; PubMed-9476126;
RA Hellwege J., Skerka C., Zipfel P.F.;
RT "Biochemical and functional characterization of the factor-H-related
RT protein 4 (FHR-4).";
RL Immunopharmacology 38:149-157(1997).
CC -!- FUNCTION: INVOLVED IN COMPLEMENT REGULATION. CAN ASSOCIATE WITH
CC LIPOPROTEINS AND MAY PLAY A ROLE IN LIPID METABOLISM.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- PTM: GLYCOSYLATED.
CC -!- SIMILARITY: CONTAINS 5 SUSHI (SCR) DOMAINS.
CC -!- SIMILARITY: STRONG, TO FACTOR H.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC

```



modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL: X98337; CAA66980.1; ..  
 DR EMBL: AFI90816; AAF05951.1; ..  
 DR HSSP: P10998; 1WVD.  
 DR MIM: 605337; ..  
 DR InterPro: IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam: PF00084; sushi; 5.  
 DR SMART: SM00032; CCP; 4.  
 KW Repeat: Glycoprotein; Sushi; Signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 331 COMPLEMENT FACTOR H-RELATED PROTEIN 4.  
 FT DOMAIN 23 85 SUSHI 1.  
 FT DOMAIN 87 146 SUSHI 2.  
 FT DOMAIN 148 205 SUSHI 3.  
 FT DOMAIN 210 266 SUSHI 4.  
 FT DOMAIN 268 331 SUSHI 5.  
 FT DISULFID 24 73 BY SIMILARITY.  
 FT DISULFID 56 84 BY SIMILARITY.  
 FT DISULFID 88 134 BY SIMILARITY.  
 FT DISULFID 117 145 BY SIMILARITY.  
 FT DISULFID 149 193 BY SIMILARITY.  
 FT DISULFID 176 204 BY SIMILARITY.  
 FT DISULFID 211 254 BY SIMILARITY.  
 FT DISULFID 240 265 BY SIMILARITY.  
 FT DISULFID 269 320 BY SIMILARITY.  
 FT DISULFID 303 330 BY SIMILARITY.  
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 331 AA; 37325 MW; 5A0D04AB84B841424 CRC64;

Query Match 21.4%; Score 681; DB 1; Length 331;  
 Best Local Similarity 28.3%; Pred. No. 5e-43;  
 Matches 163; Conservative 48; Mismatches 114; Indels 250; Gaps 11;

QY 1 MLLFVSLISWTVGG-EGTLCDFPKIHGHFLYDEEDYNPFSQVPTGVFVYSCRYNF 59  
 DB 1 MLLINVLNWSKANGQVQKCDPFEIOHGLYKSLRLFPAAAGSYSYCDQNF 60  
 QY 60 VSPKSFWRITTEGWSPTKCLRMCSFP--FVKNHSHSEGLIHLEGTQIIICNTG 117  
 DB 61 VTPSGSYDVIHCTQDGLPVLCLRTCKSDIEIENGFTSESSIVILNKEIQYCKPG 120  
 QY 118 YSLQ--NNEKNISCVRGWSTPPIG-SFTKGCHVPILNVAQPKESYKVGDLVLF 174  
 DB 121 YATADNGSGSITCLQNGWSAQPICTIKF---CDMPVFE-NSRAKSNMGKFLKLDLDYE 175  
 QY 175 CRKNLIRVGSVQCYQFGNSPFPPTCKGVRSQVPPQLSNGEVKEIRKEEYGHNEVVE 234  
 DB 176 C----- 176  
 QY 235 YDCNPFIINGPKKIQCVGDEWTLTPTCVQVKTGYPILEYGVQVPPYQHGVSVE 294  
 DB 177 YD-----GY--EISYG----- 185  
 QY 295 VNCNEYAMTGNWITINGIWTLPVCVATHQLKCKKIAGVNIKTLKLSGREFNNSR 354  
 DB 186 ----- 185  
 QY 355 IRYRCSDFRYRHSVCINGKNPEVDCTEKREQPCPPPPQIPNAQNMTTVNYQDGKVA 414  
 DB 186 -----NTGS----- 190  
 QY 415 VLCKENYLLPEAKIVCKGRWQSLPRCVSTAYCGPPPSINNGDTTSFPLSVYPPGSTV 474  
 DB 191 -----IVCGEGWGHFFPCYNSSEKCGPPPLISNGDTSFLLKVVVPSRV 236  
 QY 475 TYRQSFYKLGQSVTVTCRNKQWSEPPRCCLDPCVVSEENMKNKNIQLKWRDGLKYARTG 534  
 DB 475 :||||:|||| ||| : |||||: ||:::||||||| ||| : |||||

Db 237 EYQCQSYELQGSNYVTCSNGWSEPPRCIHPCITTEENMKNKNIQLKGRSDIKYAKTG 296  
 QY 535 DAVEFOCKFPHKAMISSPPFRAICQCKGFEYPICE 569  
 Db 297 DTIEFMCKLGYNTANSTVLSFQAVCREGIVETPRCE 331  
 RESULT 10  
 CR2\_HUMAN  
 ID CR2\_HUMAN STANDARD; PRT; 1033 AA.  
 AC P20023; Q13866; Q14212;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Complement receptor type 2 precursor (Cr2) (Complement C3d receptor)  
 DE (Epstein-Barr virus receptor) (EBV receptor) (CD21 antigen).  
 GN CR2 OR C3DR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89123277; PubMed=2563370;  
 RA Fujisaku A., Harley J.B., Frank M.B., Gruner B.A., Frazier B.,  
 RA Holers V.M.;  
 RT "Genomic organization and polymorphisms of the human C3d/Epstein-Barr  
 RT virus receptor";  
 RL J. Biol. Chem. 264:2118-2125(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE-B-cell;  
 RC MEDLINE=88171282; PubMed=2832506;  
 RA Weis J.J., Toothaker L.E., Smith J.A., Weis J.H., Fearon D.T.;  
 RT "Structure of the human B lymphocyte receptor for C3d and the  
 RT Epstein-Barr virus and relatedness to other members of the family  
 RT C3/C4 binding proteins";  
 RL J. Exp. Med. 167:1047-1066(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM C).  
 RX MEDLINE=88097454; PubMed=2827171;  
 RA Moore M., Cooper N., Rack B., Nemerow G.;  
 RT "Molecular cloning of the cDNA encoding the Epstein-Barr virus C3d  
 RT receptor (complement receptor type 2) of human B lymphocytes";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:9194-9198(1987).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORMS A; C AND D).  
 RX MEDLINE=99165386; PubMed=10068037;  
 RA Barel M., Balbo M., Frade R.;  
 RT "Evidence for a new transcript of the Epstein-Barr virus/C3d receptor  
 RT (CR2, CD21) which is due to alternative exon usage";  
 RL Mol. Immunol. 35:1025-1031(1998).  
 RN [5]  
 RP SEQUENCE OF 226-233; 256-267; 332-341; 567-677 AND 898-908.  
 RX MEDLINE=86287311; PubMed=3016712;  
 RA Weis J.J., Fearon D.T., Klickstein L.B., Wong W.W., Richards S.A.,  
 RA de Bruyn Kops A., Smith J.A., Weis J.H.;  
 RT "Identification of a partial cDNA clone for the C3d/Epstein-Barr  
 RT virus receptor of human B lymphocytes: homology with the receptor for  
 RT fragments C3b and C4b of the third and fourth components of  
 RT complement";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5639-5643(1986).  
 RN [6]  
 RP SEQUENCE OF 492-556 FROM N.A. (ISOFORM B).  
 RX MEDLINE=93294286; PubMed=8390533;  
 RA Sinha S.K., Todd S.C., Hedrick J.A., Speiser C.L., Lambris J.D.,  
 RA Tsoukas C.D.;  
 RT "Characterization of the EBV/C3d receptor on the human Jurkat T cell  
 RT line: evidence for a novel transcript";  
 RL J. Immunol. 150:5311-5320(1993).  
 CC - FUNCTION: Receptor for complement C3d and for the Epstein-Barr  
 CC virus on human B-cells and T-cells. Participates in B lymphocytes  
 CC activation.



```

QY 135 STPICGFTGECVHPILF-----ANV-----DAQP-----160
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
266 TKMPVCEELFCPSPPILNGRHIGNSLANVSGSIVTYTCDPDEEGVNFILIGESTLRC 325
QY 161 -----KESYKVGDLVLFSCRKMLIR 181
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
326 TVDSQKGTGSGPAPRGELSTSAVQCPHPQILGRMVSGQDRYTYNDTVIFACMGFTL 385
QY 182 VGSVVOCYQFG-WSPNFPCTKGVRSCGPPQLSNGEVRKIRKEEYGHNVVBYDCNPN 240
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
386 KGSQIRCNAGTWEPSNPVE---KECOAPPNLLNGKEDRHVRDPGTSIKYSCNPG 442
QY 241 FIINGPKKIQCVDGEMT-TLPTCQVQKTC-----GYIPELEYGYVOPSV-----284
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
443 YVLVGEESIOCTSEGVMTPPVQC--KVAACEATGROLLTRPQHOFVRPVDVNSCGEGYK 500
QY 285 -----PP-----YQHGVSVVENCN--299
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
501 LSGSVQECQGTIPWFWEIRLCKEITCPPPPVIYNGAHTGSLSEDFPYGTVTTCNPGP 560
QY 300 ----EYAMIGNMITCIN-----GIWT--ELPMCVATHOLKCRKCIAGVNIKTLLKLSGKE-348
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
561 ERGVEFSLIGESTIRCTSDERGTWSPAPLCKLSLLAVQC--SHVHIANGYKISGKEA 618
QY 349 -FNHNSRIYRCSDIFRYRHVCINGK-----WNPEVDCTEK-----REQF-----388
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
619 PFYNDTVTFKCYSGFTLKSSQIRCKRADNTWDPEIPVCEKETQCHVRQSLQELPAGSRV 678
QY 389 -----CPTPPQIPNAQNM--TITVNY 407
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
679 ELVNTSCODGVOLGHAQVQMCDAENGIFWKIPLCKVHCHPPPIVINGKHTGMAENF 738
QY 408 ODGEKVAVLCKENYLLPEAKIVCKD-----GRWQS-LPRCVES-----445
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
739 LYGVNEVYECQGVYLLGKELKQCRSDSKHGSGSPQLRSPPTVTRCPNPEVKHGYK 798
QY 446 ----TAY-----CGPPPSIN 456
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
799 LNKTHSAYSHNDIVVDCNPGFIMNGSRVIRCHTDNTWVPGVPTCIKKAFTGCPPPPKTP 858
QY 457 NCDTSPFLSVYPGSTVYRCQSFYKLOGSVTVTCRNK-QWSEP-PRCLDPCVVSEENM 514
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
859 NGNHTGGIARFSPGMSTLYSCDQGYLLVGEALLCTHEGTWSQAPAPCKEYVNCSSPADM 918
QY 515 MNNTQLKWRNDGKLYAKTGDV-----EFQCKFPHKAMISPPPRFAICQE 560
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
919 --DGIO-KGLEPRKMY-QYGAVVTLECEDGYMLGCSPOSQCSQSDHQ----WNPPL-AVCRS 970
QY 561 GKFEYPIC 568
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
971 RSLAPVLC 978

RESULT 11
CRL_HUMAN STANDARD; PRT; 2039 AA.
AC P17927;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Complement receptor type 1 precursor (C3b/C4b receptor) (CD35
DE antigen).
DE CRI OR C3BR.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89035992; PubMed=2972794;
RA Klickstein L.B., Barton T.J., Miletic V., Rabson L.D., Smith J.A.,
RA Fearon D.T.;
RT *Identification of distinct C3b and C4b recognition sites in the

```

```

RT human C3b/C4b receptor (CRI, CD35) by deletion mutagenesis.*;
RL J. Exp. Med. 168:1699-1717(1988).
RN [2]
RP SEQUENCE OF 503-2039 FROM N.A.
RX MEDLINE=87168191; PubMed=2951479;
RA Klickstein L.B., Wong W.W., Smith J.A., Weis J.H., Wilson J.G.,
RA Fearon D.T.;
RT *Human C3b/C4b receptor (CRI). Demonstration of long homologous
RT repeating domains that are composed of the short consensus repeats
RT characteristics of C3/C4 binding proteins.*;
RL J. Exp. Med. 165:1095-1112(1987).
RN [3]
RP SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.
RX MEDLINE=86067975; PubMed=2933745;
RA Wong W.W., Klickstein L.B., Smith J.A., Weis J.H., Fearon D.T.;
RT *Identification of a partial cDNA clone for the human receptor for
RT complement fragments C3b/C4b.*;
RL Proc. Natl. Acad. Sci. U.S.A. 82:7711-7715(1985).
CC -!- FUNCTION: CRI, PRESENT ON ERYTHROCYTES, LEUKOCYTES, GLOMERULAR
CC PODOCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS. MEDIATES THE
CC BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE
CC ACTIVATED COMPLEMENT.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- POLYMORPHISM: CRI IS RESPONSIBLE FOR THE KNOPS BLOOD GROUP SYSTEM.
CC -!- MISCELLANEOUS: SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A
CC LONG HOMOLOGOUS REPEAT (LHR). THE N-TERMINAL TWO SRCS OF LHR-A
CC CONTAINED A SITE DETERMINING C4 SPECIFICITY, AND THE N-TERMINAL
CC TWO SRCS OF LHR-B AND -C EACH HAD A SITE DETERMINING C3
CC SPECIFICITY.
CC -!- MISCELLANEOUS: THIS IS THE SEQUENCE OF THE F ALLOTYPE OF CRI.
CC -!- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION
CC (RCA) FAMILY.
CC -!- SIMILARITY: CONTAINS 30 SUSHI (SCR) DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11569; AAA52297.1; -
DR EMBL; M11617; AAA52298.1; -
DR EMBL; M11618; AAA52299.1; -
DR EMBL; Y00816; CAA68755.1; -
DR EMBL; X05309; CAA28933.1; -
DR PIR; A28507; A28507.
DR PIR; A24748; A24748.
DR PIR; B24748; B24748.
DR PIR; C24748; C24748.
DR PIR; S03843; S03843.
DR HSSP; P08603; IHFI.
DR Gene; HGNC:2334; CRI.
DR MIM; 120620; -
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 30.
DR SMART; SM00032; CCP; 30.
KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW Receptor; Sushi; Blood group antigen.
FT SIGNAL 1 41
FT CHAIN 42 2039
FT DOMAIN 42 1971
FT TRANSMEM 1972 1996
FT DOMAIN 1997 2039
FT MOD_RES 42 42
FT DOMAIN 42 100
FT DOMAIN 103 162
FT DOMAIN 165 233
FT DOMAIN 237 294
FT DOMAIN 296 354
FT DOMAIN 357 417
FT COMPLEMENT RECEPTOR TYPE 1.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT PYRROLIDONE CARBOXYLIC ACID (POTENTIAL).
FT SUSHI A1.
FT SUSHI A2.
FT SUSHI A3.
FT SUSHI A4.
FT SUSHI A5.
FT SUSHI A6.

```

```
FT DOMAIN 420 488 SUSHI A7.
FT DOMAIN 492 550 SUSHI B1.
FT DOMAIN 553 612 SUSHI B2.
FT DOMAIN 615 683 SUSHI B3.
FT DOMAIN 687 744 SUSHI B4.
FT DOMAIN 746 804 SUSHI B5.
FT DOMAIN 807 867 SUSHI B6.
FT DOMAIN 870 938 SUSHI B7.
FT DOMAIN 942 1000 SUSHI C1.
FT DOMAIN 1003 1062 SUSHI C2.
FT DOMAIN 1065 1133 SUSHI C3.
FT DOMAIN 1137 1194 SUSHI C4.
FT DOMAIN 1196 1254 SUSHI C5.
FT DOMAIN 1257 1317 SUSHI C6.
FT DOMAIN 1320 1388 SUSHI C7.
FT DOMAIN 1395 1453 SUSHI D1.
FT DOMAIN 1456 1515 SUSHI D2.
FT DOMAIN 1518 1586 SUSHI D3.
FT DOMAIN 1590 1647 SUSHI D4.
FT DOMAIN 1649 1707 SUSHI D5.
FT DOMAIN 1710 1770 SUSHI D6.
FT DOMAIN 1773 1841 SUSHI D7.
FT DOMAIN 1847 1905 SUSHI E1.
FT DOMAIN 1908 1966 SUSHI E2.
FT DISULFID 43 86 BY SIMILARITY.
FT DISULFID 73 99 BY SIMILARITY.
FT DISULFID 104 145 BY SIMILARITY.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 238 280 BY SIMILARITY.
FT DISULFID 266 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 358 400 BY SIMILARITY.
FT DISULFID 386 416 BY SIMILARITY.
FT DISULFID 421 470 BY SIMILARITY.
FT DISULFID 450 487 BY SIMILARITY.
FT DISULFID 493 536 BY SIMILARITY.
FT DISULFID 523 549 BY SIMILARITY.
FT DISULFID 554 595 BY SIMILARITY.
FT DISULFID 581 611 BY SIMILARITY.
FT DISULFID 616 665 BY SIMILARITY.
FT DISULFID 645 682 BY SIMILARITY.
FT DISULFID 688 730 BY SIMILARITY.
FT DISULFID 716 743 BY SIMILARITY.
FT DISULFID 747 790 BY SIMILARITY.
FT DISULFID 776 803 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 866 BY SIMILARITY.
FT DISULFID 871 920 BY SIMILARITY.
FT DISULFID 900 937 BY SIMILARITY.
FT DISULFID 943 986 BY SIMILARITY.
FT DISULFID 973 999 BY SIMILARITY.
FT DISULFID 1004 1045 BY SIMILARITY.
FT DISULFID 1031 1061 BY SIMILARITY.
FT DISULFID 1066 1115 BY SIMILARITY.
FT DISULFID 1095 1132 BY SIMILARITY.
FT DISULFID 1138 1180 BY SIMILARITY.
FT DISULFID 1166 1193 BY SIMILARITY.
FT DISULFID 1197 1240 BY SIMILARITY.
FT DISULFID 1226 1253 BY SIMILARITY.
FT DISULFID 1258 1300 BY SIMILARITY.
FT DISULFID 1286 1316 BY SIMILARITY.
FT DISULFID 1321 1370 BY SIMILARITY.
FT DISULFID 1350 1387 BY SIMILARITY.
FT DISULFID 1396 1439 BY SIMILARITY.
FT DISULFID 1426 1452 BY SIMILARITY.
FT DISULFID 1457 1490 BY SIMILARITY.
FT DISULFID 1484 1514 BY SIMILARITY.
FT DISULFID 1519 1568 BY SIMILARITY.
FT DISULFID 1548 1585 BY SIMILARITY.
FT DISULFID 1591 1633 BY SIMILARITY.

FT DISULFID 1619 1646 BY SIMILARITY.
FT DISULFID 1650 1693 BY SIMILARITY.
FT DISULFID 1679 1706 BY SIMILARITY.
FT DISULFID 1711 1753 BY SIMILARITY.
FT DISULFID 1739 1769 BY SIMILARITY.
FT DISULFID 1774 1823 BY SIMILARITY.
FT DISULFID 1803 1840 BY SIMILARITY.
FT DISULFID 1848 1891 BY SIMILARITY.
FT DISULFID 1877 1904 BY SIMILARITY.
FT DISULFID 1909 1952 BY SIMILARITY.
FT DISULFID 1938 1965 BY SIMILARITY.
FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 509 509 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 860 860 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 959 959 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1028 1028 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1152 1152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1310 1310 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1481 1481 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1504 1504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1540 1540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1605 1605 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1763 1763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1908 1908 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2039 AA; 223589 MW; B2FD29B6AD3C5EB7 CRC64;

Query Match 12.1%; Score 385; DB 1; Length 2039;
Best Local Similarity 25.4%; Pred No. 2.3e-20;
Matches 142; Conservative 78; Mismatches 204; Indels 136; Gaps 33;

QY 37 EDYNPFSQVPT----GEVFIYSCYENF---VSPSKSFWTTRITCTEENST-PRCLRMCS 88
DB 1462 EPFNGMVHINTDQFGSTVNYSCNEGFRLGSPS-----TTCVSGNNVTDWKAPICE 1515
QY 89 F-----PFVKNG--HSESSGLHLBGDTVQIICNTGYSLN-----NEKNISCVERG-- 133
DB 1516 IISCEPPPTLSNGDFYSNNRTSFH-NGTVVYQCHTGPQGEQLEFELGERSIYCTSKDDQ 1574
QY 134 ---WSTPPICSTKGECHVPILKANVDAQPKESYKVGVDYKLFSCRNLRVGSDSVQCY 190
DB 1575 VGVWSSPPPRCISTNKCTAPEVENAIRVPCNRRSFFSLTEIRFCQPGFVWVGSHTVQCO 1634
QY 191 QFG-WSPNFPCTKGQVRSCGPPPOLANGEVKEIRKEEYGHNEVVEYDCNPNFTIINGPKRI 249
DB 1635 TNGRWGPKLPHCS---RVQPPPEILLGHEHTLSDHNFSPGQEVFVSCPSYDLRGAASL 1691
QY 250 QCV-DGEWT-TLPTCYEQVQKTC-GYIPELEYGVYQVSPV-PYQHGVSVEVNCNEVAMIG 305
DB 1692 HCTPQGDWSPEARCT--VKSCDDFLGQLPHGRV--LLPLNLQLGAKRVFVCDGEPRLKG 1747
QY 306 NNMITCINGIWTPLMCMVATHQLKCKIAGVNIKTLLKLSGKKEFNHNSRIRYRSCSDFRY 365
DB 1748 RS-----ASH-----CVLAG--NKAL-----1761
QY 366 RHVVCINGKNPEVDCTEKREQCPPPPOIPNNAQNMT---TVNYODGEKVAVLCKEN-- 420
DB 1762 -----WNSVPVCE--QIFCPNPPALLNGRHTGTFPGDIPY--GKEISYACDTHPD 1808
QY 421 -----YLLPEAKEIVCK-----DGRWQS-LPRCVST-AYCGPPPSINNGDTTSFPPLSVVP 469
DB 1809 RGMTFNLIGESSIRCTSDPQGNVWSSPAPRCELSVPAACPHPKPQNGHYIGGHRVSLYL 1868
QY 470 PGSTVTRYCQSFYKLGQSVTVTCRNKO-WSE-----PPRCLDPCVVSENNKNKIQLK 522
DB 1869 PGWTSYTCDPGLVVGKGFIFCTDQGIWSQLDHYCKEYVNCSPFLF-----MNGISKELE 1923
```



FT	CONFLICT	520	520	P -> A (IN REF. 2).
FT	CONFLICT	962	963	MISSING (IN REF. 4).
SQ	SEQUENCE	1025	AA; 112994	MW; 19E518B9A0273694 CRC64;
Query Match				
Best Local Similarity		24.28;	Pred. No. 1.7e-20;	
Matches		158;	Conservative	76; Mismatches 268; Indels 150; Gaps 36;
QY	41	PFSQVP--TGEVFYISC--EYNFVSPSKSEFWTRITCTEEG-----WSPTPKCLR-----	85	
DB	26	PYISLPVPCVTLRYTCSPSYRLIGKAIF-----CISENQVHATWDKAPICESYVKTI	80	
QY	86	MCSEFPVKNGH-SESSGLIHLEGTVOIICNTGYSLQNNKNISCVR---GWSPTPIC-144		
DB	81	SCSDPIVPGFMFMKSGKAPRHRGDSVFTTCKANFTMKS-KTMVQCANMGMPTALPVC	139	
QY	141	SFTKGEC-HVPILEANYDAQPKESYKVGDLVLFSCRNLRVCGSDSVQCYQFG-WSPNF	198	
DB	140	SDFLECPSLPTTHNGHHTQHQDQFVAGUSVTVSCBPGYLLTGKTKIKCLSSGDWDGVI	199	
QY	199	PTCKGQVRSCGPPPOLSNGBEVKEIRKEEYGHNEVVEYDCNPFNTINGPKIKQYDGE---	255	
DB	200	PTCK--EAOCEHCGKFPNGQVKPELSLQVG--ITVYFSCNEGVLQGOQSSQCVIVQKA	255	
QY	256	-WTTLPTCVDQVKTGCIPEL-----EYGYVQPSVPPIYOHGVSVEVNCN-----EYAMI	304	
DB	256	IWTKKPVCKEIL--CPPPPPVRNGSHGTSESNVP---YGSTVITYCDPSPEKGVSTLI	310	
QY	305	GNNMITCI-----NGIWT-ELPMCVATHOLKRC--KIAGVNIKTLLKSGKFNHNSRI	355	
DB	311	GKTINGTGSQKTGWSGPAPYCVLSTSAVLCLQPKIKRQILSIUDKS---YSYNDTV	367	
QY	356	RYRCSDFIRYHSVCI-----NGKNWPEVDCTEKREQFCPPPIPAQNNMTT-TVNYODG	410	
DB	368	AFSECPGFTLGNRSIRCNAGHTWEPVPVCEG--COAPPKINGKEDSYLLNFDPG	424	
QY	411	EKVAVILCKENYLLPEAKEIVC-KDGRWOSL-PRC-----	442	
DB	425	TSIRYSCDPGYLLVGEDTIHCTPEGKWTPITPOCTVAECKPVGPHLFKRQNFIRTAVN	484	
QY	443	-----VESTAY-----CGPPPSINGDDTSPFLSVYPCGSTV	474	
DB	485	SSCDEGFQLSESAQLCOGTPWPIEIRLCKEITCPPPVVINGHTHTWSSSESDVPYGTVV	544	
QY	475	TYRQCS-----FYKLGVSVTVCRNK-----QWSEP-PRC--LDPCVWSENNKNNIQ	520	
DB	545	TYMCPGPBEGVKFLIGEOTIHCTSDSRGRGWSWSPAPLKLISLPAVQCTDHHVENGVR	604	
QY	521	LKWRNDGKLYAKTGDAVEFCQKPPHKAMITSSPPFRAICQEGKF---EYPICE	569	
DB	605	L---TDNKAPYFTNDSVMFKC---DDGYILSGSSQIRCKANNTWDEPKPLK	650	

RESULT 13

C4BP_RAT	C4BP_RAT	STANDARD;	PRT;	558 AA.
ID	C4BP_RAT			
AC	Q63514;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	C4b-binding protein alpha chain precursor (C4bp).			
DE	C4BPA.			
GN	Rattus norvegicus (Rat).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN:Sprague-Dawley; TISSUE:Liver;			
RX	MEDLINE-97166082; PubMed-9013975;			
RA	Hillarp A., Wiklund H., Thern A., Dahlback B.;			
RT	"Molecular cloning of rat C4b binding protein alpha- and beta-chains:			
RT	structural and functional relationships among human, bovine, rabbit,			

```
mouse, and rat proteins.";
```

```
RL J. Immunol. 158:1315-1323(1997).
```

```
-!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
```

```
CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
```

```
CC (C3BINA), WHICH THEN HYDROLYSES THE COMPLEMENT FRAGMENT C4B. IT
```

```
CC ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX.(C3. IT
```

```
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
```

```
CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S
```

```
CC AND WITH SERUM AMYLOID P COMPONENT.
```

```
CC -!- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.
```

```
CC -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
```

```
CC -!- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.
```

```
-----
```

```
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outpatient-
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce)
```

```
CC or send an email to licensee@isb-sib.ch). 
```

```
-----
```

```
EMBL; Z50051; CAA90391.1; -. 
```

```
HSPG; P10998; LWVD.
```

```
InterPro; IPR000436; Sushi_SCR_CCP.
```

```
Pfam; PF00084; sushi; 8.
```

```
SMART; SM00032; CCP; 8.
```

```
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
```

```
SIGNAL      1   13
```

```
BV SIMILARITY.
```

```
FT FT     CHAIN       14    558
```

```
FT FT     DOMAIN      14    73
```

```
FT FT     DOMAIN      76   135
```

```
FT FT     DOMAIN     138   200
```

```
FT FT     DOMAIN     203   259
```

```
FT FT     DOMAIN     262   325
```

```
FT FT     DOMAIN     328   387
```

```
FT FT     DOMAIN     389   444
```

```
FT FT     DOMAIN     446   502
```

```
FT FT     DISULFID     15    60
```

```
FT FT     DISULFID     45    72
```

```
FT FT     DISULFID     77   118
```

```
FT FT     DISULFID    104   134
```

```
FT FT     DISULFID    139   182
```

```
FT FT     DISULFID    168   199
```

```
FT FT     DISULFID    204   246
```

```
FT FT     DISULFID    232   258
```

```
FT FT     DISULFID    263   312
```

```
FT FT     DISULFID    296   324
```

```
FT FT     DISULFID    328   351
```

```
FT FT     DISULFID    329   373
```

```
FT FT     DISULFID    363   386
```

```
FT FT     DISULFID    390   431
```

```
FT FT     DISULFID    417   443
```

```
FT FT     DISULFID    447   488
```

```
FT FT     DISULFID    474   501
```

```
FT FT     DISULFID    509   509
```

```
FT FT     DISULFID    521   521
```

```
FT FT     CARBOHYD     31    31
```

```
FT FT     CARBOHYD    177   177
```

```
FT FT     CARBOHYD    186   186
```

```
FT FT     CARBOHYD    469   469
```

```
FT FT     CARBOHYD    491   491
```

```
SQ SEQUENCE          558 AA; 62266 MW; 592FC667BD1E5FF CRC64;
```

```
Query Match              10.7%; Score 339.5; DB 1; Length 558;
```

```
Best Local Similarity    24.9%; Pred.No. 1.le-17;
```

```
Mates 126; Conservative 75; Mismatches 207; Indels 99; Gaps 30;
```

```
QY        109 TVQIICNTGYSLQNNEKNCISVERG-WSTPPICSFKGECHVP--ILEANYDAOPKKESY 165
```

```
Dd         | :| |||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
```

```
           40 TLRYNCRPGYSWASSQSGLCKPLGWQNIATC-VVKSCRNPFGLQNGKEVV---KTDF  94
```

```
QY        166 KYGVDLKFSCRKNLIIRVGSDSVQCYYFG----WSFNPTCTKGVRSGCGPQLNSPEYKE 221
```

```
           | ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| :
```

Db 95 LFGSOLFSCSEGYILIGSTSYCEIOGKGVSWSDPLPEC--VIAKCGMPDINSNGKHG 152  
QY 222 IRKEEYGNVEVDCNENFIINGPKKIQ-----VDGEWTLTPCQEYKTCGYIPELE 276  
Db 153 REEEFTYRSSVTKCDPDTLLGNASTCTVNNKTVGVWSPSPPTCERI-ICPW-PKVL 210  
QY 277 YGYVQPSVP-PYOHGVSVEVNCNRYAMIGNMILTC-INGIWTLPVCMVATHOLKRC-KI 333  
Db 211 HGTINSQPKHYKYKDSYFVCKGVLRGSGVHCEADGWSVPVPC-----ELNSCTDI 266  
QY 334 AGVNIKTLLKSGKFNENSRIR-----YRCSDFIRYHSVING----- 373  
Db 267 PDIPNAALIT-----SPRPRKEDYFVGVTLRY-----ICRPGYEPATQPMVTCQKD 315  
QY 374 -KNWPEVDCTEKREQFCPPPPQEPNAQ-----NMVTVVYQDGEKVAVLCKENYLLPE 425  
Db 316 LSWMLRCKE-----ICCPVDPKSVRVIOHEKAHPDNDCTFFGDEVSVYTCNDIML-- 369  
QY 426 AKEIVCK-DGRWQ-SLPRCVSTAYCGPPPPINDDTTSFPLSVVPPGTVTYRCQSFYK 483  
Db 370 --TATCKSDGTWHPRTSPCHOS---CDFPPIAHGRTK---SSSYVVRTQVYCEEGYR 422  
QY 484 LQGSVTVTCRNKQWS-EPRLDPC---VWSEENKNNIQLKWRNDGKLYAKTGDAVEF 539  
Db 423 LVGEATISQWYSOWTPAAPQCKALCRKPEIGNGVLSTNKDQ-----YVET-ENVTI 472  
QY 540 OCKEPHKAMISSPPRAICQEGKFEY 566  
Db 473 QCDSCF-VMLGSQSI--TCSNGWTWP 496

## RESULT 14

C4BP\_BOVIN  
ID C4BP\_BOVIN STANDARD; PRT: 610 AA.  
AC Q28065;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE C4b-binding protein alpha chain precursor (C4bp).  
GN C4BPA.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=95015909; Pubmed=7930621;  
RT Hillarp A., Thern A., Dahlbaech B.;  
RT "Bovine C4b binding protein. Molecular cloning of the alpha- and  
RT beta-chains provides structural background for lack of complex  
RT formation with protein S";  
RL J. Immunol. 153:4190-4199(1994).  
CC -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT  
CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR  
CC (C3BI), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT  
CC ALSO ACCELERATES THE DEGRADATION OF THE C4B/C2A COMPLEX (C3  
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A' ALPHA  
CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.  
CC -!- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS  
CC (BY SIMILARITY).  
CC -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.  
CC -!- SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; Z31693; CAA83498.1; -.  
DR HSP: P10998; LVVD.  
DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
DR Pfam; PF00084; sushi; 8.  
DR SMART; SM00032; CCP; 8.  
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.  
FT SIGNAL 1 48  
FT CHAIN 49 610  
FT DOMAIN 49 108  
FT SUSHI 1.  
FT DOMAIN 111 170  
FT SUSHI 2.  
FT DOMAIN 173 235  
FT SUSHI 3.  
FT DOMAIN 238 295  
FT SUSHI 4.  
FT DOMAIN 298 363  
FT SUSHI 5.  
FT DOMAIN 366 426  
FT SUSHI 6.  
FT DOMAIN 428 484  
FT SUSHI 7.  
FT DOMAIN 486 542  
FT DISULFID 50 95  
FT BY SIMILARITY.  
FT DISULFID 80 107  
FT BY SIMILARITY.  
FT DISULFID 112 153  
FT BY SIMILARITY.  
FT DISULFID 139 169  
FT BY SIMILARITY.  
FT DISULFID 174 217  
FT BY SIMILARITY.  
FT DISULFID 203 234  
FT BY SIMILARITY.  
FT DISULFID 239 281  
FT BY SIMILARITY.  
FT DISULFID 267 294  
FT BY SIMILARITY.  
FT DISULFID 299 350  
FT BY SIMILARITY.  
FT DISULFID 334 362  
FT BY SIMILARITY.  
FT DISULFID 366 390  
FT BY SIMILARITY.  
FT DISULFID 367 412  
FT BY SIMILARITY.  
FT DISULFID 429 471  
FT BY SIMILARITY.  
FT DISULFID 457 483  
FT BY SIMILARITY.  
FT DISULFID 487 528  
FT BY SIMILARITY.  
FT DISULFID 514 541  
FT BY SIMILARITY.  
FT DISULFID 549 549  
FT DISULFID 561 561  
FT CARBOHYD 66 66  
FT CARBOHYD 221 221  
FT CARBOHYD 525 525  
FT CARBOHYD 602 602  
SQ SEQUENCE 610 AA; 58886 MW; D806B2708BA06B58 CRC64;  
Query Match 10.3%; Score 328; DB 1; Length 610;  
Best Local Similarity 22.8%; Pred. No. 8.7e-17;  
Matches 117; Conservative 76; Mismatches 218; Indels 102; Gaps 26;  
QY 107 GDTVQLICNTGYSLONNEKNISC-VERGWSPTTICSTFKGECVHP--ILEANDVQPKKE 163  
Db 73 GTTLRYTCRPGYRISRRKNFLICDGTDMWYKEFC--VKKRCENPGLLNGQVIV---KT 127  
QY 164 SYKVGDLVKFSCRKNLIRVSDSVQY---QFGWSPNFPCTCKGVRSGCPPPQLSNGEV 219  
Db 128 DYSFGESEIEFSCSEGYVLIGSANSYCOLQDKGVWSDPLQPC--LIANCEPPTISNGRH 185  
QY 220 KEIRKEEYGHNEVEYDCNPNFIINGPKKIQ-----VDGEWTLTPCQEYKTCGYIPE 274  
Db 186 NGDEDFYTYGSSVYTSYCDRDFSMLGKASISCRVENKTIQVWSPSPSKKV-IC----- 239  
QY 275 LEYGVVQPSV-----PEYQHGVSVEVNCNRYAMIGNMILTC-INGIWTLPVCMV 323  
Db 240 -----VQPVWDGKITSGFGPIYTYQQSIVIVACNKGFRLEGDSLTHCEADNSWNPDP 291  
QY 324 ATHQLKRCIAGVNIKTLLKLSGKFNENSRIRYRCSDI-----FRYRHS----- 368  
Db 292 PTCELNGC-----LGLPHIPHALWERYDQTQTEQQQYVDIGFVLSYKCHGYKPEPDGPT 347  
QY 369 -VC-INGKWNPEVDCTEKREQFCPPPPQIPNQNMT-----TTVNVQDEKAVLCK 418  
Db 348 VTCQNLWSPVYIECKE-----VCCPEPNLNNYTSITLHRRPSTSTHCTYISGDKISYEC 403  
QY 419 ENYLLPEAKEIVC-KDGRWQSLPRCVSTAYCGPPSPINNGD---TTSFPLSVYPPGSTV 474  
Db 404 SKYMF-----DALCTKHGTWS---PRTPECRDCKSPPVIAHGOHKVYSKF---FTFDHQA 453

QY 475 TYRCSPYKLOGSVTVTCRNKQKSEP-PRCLDPCVSEENWKNKNIQKWNCKLYAKT 533  
Db 454 VYEDKGYLVGAKELSGSGSPVAPQCKALCKPKEIYGRLSVE-----KVRYVE 506  
QY 534 GDAVEFOCKPFPKAMISPPFRATCQGGKPEYP 566  
Db 507 PEITIOCESGYVVGSE---NITCSEDRTWYP 536

RESULT 15  
ID C4BP\_HUMAN STANDARD; PRT; 597 AA.  
AC P04003;  
DT 23-OCT-1986 (Rel. 02, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE C4b-binding protein alpha chain precursor (C4bp) (Proline-rich  
DE protein) (PRP).  
GN C4BPA OR C4BP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=30073699; PubMed=2590215;  
RA Matsuguchi T., Okamura S., Aso T., Sata T., Niho Y.;  
RT "Molecular cloning of the cDNA coding for proline-rich protein (PRP):  
RT Identity of PRP as C4b-binding protein.";  
RL Biochem. Biophys. Res. Commun. 165:138-144(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91113199; PubMed=1989602;  
RA Aso T., Okamura S., Matsuguchi T., Sakamoto N., Sata T., Niho Y.;  
RT "Genomic organization of the alpha chain of the human C4b-binding  
RT protein gene.";  
RL Biochem. Biophys. Res. Commun. 174:222-227(1991).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RA Strausberg R.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 9-81 FROM N.A.  
RX MEDLINE=88242821; PubMed=3378624;  
RA Lintin S.J., Lewin A.R., Reid K.B.M.;  
RT "Derivation of the sequence of the signal peptide in human  
RT C4b-binding protein and interspecies cross-hybridisation of the C4bp  
RT cDNA sequence.";  
RL FEBS Lett. 232:328-332(1988).  
RN [5]  
RP SEQUENCE OF 203-288 FROM N.A.  
RX MEDLINE=86301119; PubMed=3017751;  
RA Lintin S.J., Reid K.B.M.;  
RT "Studies on the structure of the human C4b-binding protein gene.";  
RL FEBS Lett. 204:77-81(1986).  
RN [6]  
RP SEQUENCE OF 80-597 FROM N.A.  
RX MEDLINE=86025405; PubMed=3840370;  
RA Chung L.P., Bentley D.R., Reid K.B.M.;  
RT "Molecular cloning and characterization of the cDNA coding for C4b-  
RT binding protein, a regulatory protein of the classical pathway of the  
RL human complement system.";  
RL Biochem. J. 230:133-141(1985).  
RN [7]  
RP SEQUENCE OF 49-88.  
RX MEDLINE=85296001; PubMed=4033666;  
RA Chung L.P., Gagnon J., Reid K.B.M.;  
RT "Amino acid sequence studies of human C4b-binding protein: N-terminal  
RT sequence analysis and alignment of the fragments produced by limited  
RT proteolysis with chymotrypsin and the peptides produced by cyanogen  
RT bromide treatment.";

Mol. Immunol. 22:427-435(1985).  
[8]  
RN ELECTRON MICROSCOPY, 3D-STRUCTURE, AND LIGAND BINDING.  
RP MEDLINE=83221615; PubMed=6222381;  
RX Dahlback B., Smith C.A., Mueller-Eberhard H.J.;  
RA "Visualization of human C4b-binding protein and its complexes with  
RT vitamin K-dependent protein S and complement protein C4b.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:3461-3465(1983).  
CC [-] FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT  
CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR  
CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4b. IT  
CC ALSO ACCELERATES THE DEGRADATION OF THE C4bC2a COMPLEX (C3  
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2a. ALPHA  
CC CHAIN BINDS C4b. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S  
CC AND WITH SERUM AMYLOID P COMPONENT.  
CC [-] SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS  
CC OF 3 POSSIBLE SORTS: A 570 kDa COMPLEX OF 7 ALPHA CHAINS AND 1  
CC BETA CHAIN, A 530 kDa HOMOPOLYMER OF ALPHA CHAINS OR A 500 kDa  
CC COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF  
CC THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE  
CC BINDING SITE FOR C4b AT THE END.  
CC [-] TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.  
CC [-] SIMILARITY: CONTAINS 8 SUSHI (SCR) DOMAINS.  
CC [-] SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.  
CC [-] CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.  
-----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M31452; AAA36507.1; -  
DR EMBL; M62486; AAA36506.1; -  
DR EMBL; M62475; AAA36506.1; JOINED.  
DR EMBL; M62476; AAA36506.1; JOINED.  
DR EMBL; M62477; AAA36506.1; JOINED.  
DR EMBL; M62478; AAA36506.1; JOINED.  
DR EMBL; M62479; AAA36506.1; JOINED.  
DR EMBL; M62480; AAA36506.1; JOINED.  
DR EMBL; M62481; AAA36506.1; JOINED.  
DR EMBL; M62482; AAA36506.1; JOINED.  
DR EMBL; M62484; AAA36506.1; JOINED.  
DR EMBL; M62485; AAA36506.1; JOINED.  
DR EMBL; BC022312; AAH22312.1; -  
DR EMBL; X07853; CAA30701.1; -  
DR EMBL; X04284; CAA51244.1; -  
DR EMBL; X04296; CAA27839.1; -  
DR EMBL; X02865; CAA26617.1; -  
DR PIR; A33568; NBHUC4.  
DR HSP; P10998; 1VVD.  
DR Genew; HGNC:1325; C4BPA.  
DR MIM; 120830; -  
DR InterPro; IPR000436; Sushi\_SCR\_OCP.  
DR Pfam; PF00084; sushi; 8.  
DR SMART; SM00032; CCP; 8.  
DR Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal;  
KW Polymorphism. 1 48  
FT SIGNAL 49 597 C4b-BINDING PROTEIN ALPHA CHAIN.  
FT CHAIN 49 109 SUSHI 1.  
FT DOMAIN 112 171 SUSHI 2.  
FT DOMAIN 174 235 SUSHI 3.  
FT DOMAIN 238 295 SUSHI 4.  
FT DOMAIN 298 361 SUSHI 5.  
FT DOMAIN 364 423 SUSHI 6.  
FT DOMAIN 425 481 SUSHI 7.  
FT DOMAIN 483 539 SUSHI 8.  
FT DISULFID 50 96 BY SIMILARITY.  
FT DISULFID 81 108 BY SIMILARITY.  
FT DISULFID 113 154 BY SIMILARITY.



FT	DISULFID	140	170	BY SIMILARITY.
FT	DISULFID	175	217	BY SIMILARITY.
FT	DISULFID	234	234	BY SIMILARITY.
FT	DISULFID	203	217	BY SIMILARITY.
FT	DISULFID	239	281	BY SIMILARITY.
FT	DISULFID	267	294	BY SIMILARITY.
FT	DISULFID	299	348	BY SIMILARITY.
FT	DISULFID	332	360	BY SIMILARITY.
FT	DISULFID	7364	7387	BY SIMILARITY.
FT	DISULFID	7365	7409	BY SIMILARITY.
FT	DISULFID	7399	422	BY SIMILARITY.
FT	DISULFID	426	468	BY SIMILARITY.
FT	DISULFID	454	480	BY SIMILARITY.
FT	DISULFID	484	525	BY SIMILARITY.
FT	DISULFID	511	538	BY SIMILARITY.
FT	DISULFID	546	546	INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT	DISULFID	558	558	INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT	CARBOHYD	221	221	N-LINKED (GLCNAC. . .).
FT	CARBOHYD	506	506	N-LINKED (GLCNAC. . .).
FT	CARBOHYD	528	528	N-LINKED (GLCNAC. . .).
FT	VARIANT	92	92	Q -> T.
FT	VARIANT	357	357	/FTID=VAR_001977.
FT	VARIANT	473	473	Y -> H.
FT	VARIANT	473	473	/FTID=VAR_001978.
FT	SEQUENCE	597 AA; 67033 MW; 67E03F2EA85A16DD CRC64;		W -> L (IN DBSNP:1801341).
				/FTid=VAR_012038.

Query Match 9.5%; Score 303; DB 1; Length 597;  
Best Local Similarity 23.3%; Pred. No. 5.8e-15;  
Matches 104; Conservative % 68; Mismatches 194; Indels 80; Gaps 26;

Qy	107	GDTVQIICNTGYSIQNNKEKINISCVERG-WSTPPICSTFKGECHVP--ILLEANVDAQPKKE	163
Db	74	GTTLTKYCLPGYVRKSHSTQTLTNSDGEWYNTFCIIYKR--CRHPGELRNGQVEI---KT	128
Qy	164	SYKVGVDLFKSCRKNLIIRVGSDSVQC---YQFGWSPNFPCTKGQVRSCTPPSPOLSNGEV	219
Db	129	DLSEFGQIEFSCSEGPFLLIGSTTSRCEVQDRGVGWSHPLQOC--EIVKCKPPPDIRNGR-	185
Qy	220	KETRKEEYGHNEVVEYDCNPNFINGPKKIQOC-----VDGEWTTLPCTCVDQVKTGCIPE	274
Db	186	HSGEENFYAGFSVYVYSCDPRFSLGHASISCTVENETIGVWRSPPTCEKI--TCRK-PD	243
Qy	275	LEYG-VYQPSVPYPQHGVSVEVNCREYAMIGNNMITC-INGIWTEPMCVAHQLRCKK	332
Db	244	VSHEGWSVGGPIYNYKDTIVFKCKQGFVLRGSSVIHCDADSKWNPSPACEPNS---C-	299
Qy	333	IAGVNIKTLKLGKKEFNHNSR-----IRYRCDIFRYRHS-----VC-INGKNWP	377
Db	300	--INLPDIPHASMETYPTRKEDVYVVGTVLRYRCHPGYKPTTDEPTTVICQKNLRWTP	356
Qy	378	EVDCTEKREQCPCPPPOIPNAQ---NMWTVN---YQDEKVAVLCKENYLLPEAKIV	430
Db	357	YQGC-----EALCCPEPKLNNGEITQHRKSRPANHCVFYGDSEIFSCHET-----SRFSAI	408
Qy	431	CK-DGRWQSI-PRCVESTAYCGPPPSINNGDTTSPPLSVYPPGST-----VTVRCQSFY	482
Db	409	COGDGWS--PRTPSCGDICNFPKPIAHGH-----YKQSSSYSPFKEELIIECDKGY	458
Qy	483	KLOGSVTVTCRNKOWSEP-PRCLDPC	507
Db	459	ILVGQAKLSCSYSHWSAPAPQCKAIC	484

Search completed: July 17, 2003, 13:51:26  
Job time : 29 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 13:49:34 ; Search time 43 Seconds  
(without alignments)  
1272.104 Million cell updates/sec

Title: US-09-805-337A-2

Perfect score: 3183

Sequence: 1 MLLLFVILISWSTVGEG.....SSPPFRAICQKGFPEICE 569

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1538	48.3	1231	1 NBHUH	complement factor
2	1524	47.9	808	2 D35069	complement factor
3	1427.5	44.8	1234	1 NBWSH	complement factor
4	1050.5	33.0	330	2 I56100	complement factor
5	927.5	29.1	270	2 I37278	complement factor
6	906.5	28.5	343	2 G35070	apolipoprotein H-r
7	861	27.0	669	2 S65551	factor H - bovine
8	783	24.6	668	2 A46013	coagulation factor
9	764.5	24.0	661	1 KFRU13	coagulation factor
10	710	22.3	331	2 A45222	complement factor
11	646	20.3	452	2 A35068	complement factor
12	476	15.0	1053	2 S46199	probable complement
13	430	13.5	1091	1 PL0009	complement C3d/Eps
14	387.5	12.2	303	2 H35068	apolipoprotein H-r
15	387.5	12.2	2014	2 I36936	complement recepto
16	383	12.0	2489	2 I73012	complement C3b/C4b
17	382	12.0	676	2 A45900	complement C3b rec
18	341.5	10.7	597	1 A43526	complement C3d/Eps
19	340.5	10.7	579	2 A56740	C4BP alpha chain p
20	339.5	10.7	558	2 S57953	sperm-egg recognit
21	328	10.3	610	1 I46001	C4BP protein alpha
22	328	10.3	610	1 I46001	C4b-binding protei
23	309.5	9.7	449	1 NBHUHS	complement factor
24	303	9.5	597	1 NBHUC4	C4b-binding protei
25	300.5	9.4	469	1 NBWSC4	C4b-binding protei
26	286.5	9.0	1797	2 T21889	hypothetical prote
27	286.5	9.0	1805	2 T21888	hypothetical prote
28	282.5	8.9	363	2 A45900	complement C3d/Eps
29	281	8.8	830	2 A30359	p-selectin precurs

## ALIGNMENTS

### RESULT 1 NEHUH

complement factor H precursor, long splice form [validated] - human  
C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 08-Dec-2000  
C:Accession: S00254; A54726; A61565; A26505; I72654; S66298  
R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.  
Biochem. J. 249, 593-602, 1988

A:Title: The complete amino acid sequence of human complement factor H.

A:Reference number: S00254; MUID:88134059; PMID:2963625

A:Accession: S00254

A:Molecule type: mRNA

A:Residues: 1-1231 <RIP>

A:Cross-references: EMBL:Y00716; NID:g31964; PIDN:CAA68704.1; PID:g31965

A:Note: 402-Tyr was also found

A:Note: parts of this sequence, including the amino and carboxyl ends of the mature p

R:Estall, J.; Schwaeble, W.; Pierich, M.; Weiss, E.H.

Eur. J. Immunol. 21, 799-802, 1991

A:Title: Human complement factor H: two factor H proteins are derived from alternativ

A:Reference number: A60238; MUID:91184292; PMID:1826264

A:Accession: A60238

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-56;1177-1231 <EST>

A:Note: only portions of this 4.3 kilobase mRNA were sequenced

R:Day, A.J.; Ripoche, J.; Lyons, A.; McIntosh, B.; Harris, T.J.R.; Sim, R.B.

Biosci. Rep. 7, 201-207, 1987

A:Title: Sequence analysis of a cDNA clone encoding the C-terminal end of human compl

A:Reference number: A54726; MUID:88025472; PMID:2889480

A:Accession: A54726

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 'DFRN',579-1231 <DAY>

A:Cross-references: GB:M17517; NID:g180497; PIDN:AAA52016.1; PID:g180498

A:Note: parts of this sequence were determined by protein sequencing

R:Ripoche, J.; Day, A.J.; Willis, A.C.; Belt, K.T.; Campbell, R.D.; Sim, R.B.

Biosci. Rep. 6, 65-72, 1986

A:Title: Partial characterization of human complement factor H by protein and cDNA se

A:Reference number: A61565; MUID:86188123; PMID:2938641

A:Accession: A61565

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 'MEGRHNLNAKI',1050-1057,'T',1059-1102 <RI2>

R:Sim, R.B.; DiScipio, R.G.

Biochem. J. 205, 285-293, 1982

A:Title: Purification and structural studies on the complement-system control protein

A:Reference number: A26505; MUID:83048213; PMID:6215918

A:Accession: A26505

A:Molecule type: protein

A:Residues: 19-20,'Q',22-29,'V',31-33,'Q',35 <SIM>

R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P.

Biochemistry 31, 3626-3634, 1992







Db 180 YEMFDEEVMCLNGNWT----- 197  
 QY 361 DIFRYRHSVCINGKWNPEVDCTEKREQCPCPPPIQIPNAQNMNTTVNYODGEKVAVLCKEN 420  
 Db 198 -----PPQ----- 200  
 QY 421 YLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDITSEPLSVYPPGSTVYRQCS 480  
 Db 201 -----CKD-----STGCGPPPPIDNGDITSFPLSVYAPASSVEYQCN 239  
 QY 481 FYKLOGSVTVTCRNKQWSEPPRCPLDPCVVSVENMNKNKNIQKWRNDGKLYAKTGDAVEFO 540  
 Db 240 LYQLEGNKRITCRNQWSEPPKCLHPCVISREIMENYIALRWTAOKOLYLRITGESAEFV 299  
 QY 541 CKFPHKAMISSPPFRAICQEGKFEYPIIC 568  
 Db 300 CKRGYRLSSRSHLTTCWDGKLEYPIC 327

## RESULT 5

I37278  
 complement factor H-related protein 2 precursor - human  
 N:Alternate names: FHR-2  
 C:Species: Homo sapiens (man)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 21-Jul-2000  
 C:Accession: I37278; I37388; S24564  
 R:Skerka, C.; Timmann, C.; Horstmann, R.D.; Zipfel, P.F.  
 J. Immunol. 148, 3313-3318, 1992  
 A:Title: Two additional human serum proteins structurally related to complement factor H  
 A:Reference number: I37278; MUID:92251200; PMID:1533657  
 A:Accession: I37278  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: mRNA  
 A:Residues: 1-270 <ZIP>  
 A:Cross-references: EMBL:X64877; NID:g30498; PIDN:CAA46096.1; PID:g30499  
 R:Skerka, C.; Moulds, J.M.; Tailon-Miller, P.; Hourcade, D.; Zipfel, P.F.  
 Immunogenetics 42, 268-274, 1995  
 A:Title: The human factor H-related gene 2 (FHR2): structure and linkage to the coagulation  
 A:Reference number: I37388; MUID:95402981; PMID:7672821  
 A:Accession: I37388  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-84, 'K', 86-143, 'S', 172-270 <SKE>  
 A:Cross-references: EMBL:X86564; NID:g1064907; PIDN:CAA60375.1; PID:g1064908  
 C:Genetics:  
 A:Gene: GDB:HFL3; FHR2  
 A:Cross-references: GDB:132658  
 A:Map position: Iq32-Iq32  
 A:Introns: 20/1; 85/1; 144/1; 171/1; 205/1  
 C:Superfamily: apolipoprotein H; complement factor H repeat homology  
 C:Keywords: duplication; extracellular protein; glycoprotein; tandem repeat  
 F:23-83/Domain: complement factor H repeat homology <FHL>  
 F:87-140/Domain: complement factor H repeat homology <FHR2>  
 F:149-203/Domain: complement factor H repeat homology <FHR3>  
 F:207-267/Domain: complement factor H repeat homology <FHR4>  
 F:126/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 29.1%; Score 927.5; DB 2; Length 270;  
 Best Local Similarity 34.6%; Pred. No. 2.5e-56;  
 Matches 198; Conservative 20; Mismatches 47; Indels 307; Gaps 3;  
 QY 1 MLLFSVLISWVSTVGEGTLCDFPKIHGGLYDEEDYNPFSQVPTGEVFFYSCYENFV 60  
 Db 1 MWLLSVLLISRISSVGGEAMPCDFPKINHGLYDEEYKYPFSQVPTGEVFFYSCYENFV 60  
 QY 61 SPKSFWRITRTCTEGSWPTPKCLRMCSFPFVKNGHSSGLIHLEGDTVQIICNTGYSL 120  
 Db 61 SPKSFWRITRTCTEGSWPTPKCLRMCSFPFVKNGHSSGLIHLEGDTVQIICNTGYSL 120  
 QY 121 QNNEKNISCVRGWSTPTPCSTKGECHVPILAEANVDAQPKKESYKGVGLVLFSCRKNLI 180  
 Db 121 QNNEKNISCVRGWSTPTPCSTKGECHVPILAEANVDAQPKKESYKGVGLVLFSCRKNLI 180  
 QY 121 QNNEKNISCVRGWSTPTPCSTKGECHVPILAEANVDAQPKKESYKGVGLVLFSCRKNLI 180  
 Db 121 QNNEKNISCVRGWSTPTPCSTKGECHVPILAEANVDAQPKKESYKGVGLVLFSCRKNLI 180

QY 181 RVGSDSVQCYQGWSPNFPCTCKGQVRSCGPPPPQLSNGEVKEIKREYGHNEVVEYDCMNP 240  
 Db 141 ----- 140  
 QY 241 FIINGPKKIQCVGDGWTTLPTCVEQVCKGYPPELEYGYVQSPVPPYQHGVSVEVNCRNE 300  
 Db 141 ----- 140  
 QY 301 YAMIGNNMTICINGIWTLPWCVATHOLKCRKIAGVNIKTLLKLSGKEFNHSRIYRCS 360  
 Db 141 ----- 140  
 QY 361 DIFRYRHSVCINGKWNPEVDCTEKREQCPCPPPIQIPNAQNMNTTVNYODGEKVAVLCKEN 420  
 Db 141 ----- 140  
 QY 421 YLLPEAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDITSEPLSVYPPGSTVYRQCS 480  
 Db 141 -----RSTISAERKCGPPPIDNGDITSFLLSVYAPASSVEYQCN 180  
 QY 481 FYKLOGSVTVTCRNKQWSEPPRCPLDPCVVSVENMNKNKNIQKWRNDGKLYAKTGDAVEFO 540  
 Db 181 LYQLEGNKRITCRNQWSEPPKCLDPCVISQIMEKYNIKLNQKLYSRGTGDIVFV 240  
 QY 541 CK---FPHKAMISSPPFRAICQEGKFEYPIIC 569  
 Db 241 CKSGYHPTK---SHSFRAMCQNGKLVYPSCE 268  
 RESULT 6  
 G35070  
 apolipoprotein H-related protein 13G1 - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 27-Jul-1990 #sequence\_revision 31-Dec-1993 #text\_change 31-Mar-2000  
 C:Accession: G35070; A35071; B35071; H35070; I35070  
 R:Vik, D.P.; Munoz-Canoves, P.; Kozono, H.; Martin, L.G.; Tack, B.F.; Chaplin, D.D.  
 J. Biol. Chem. 265, 3193-3201, 1990  
 A:Title: Identification and sequence analysis of four complement factor H-related tra  
 A:Reference number: A35070; MUID:90153969; PMID:1689298  
 A:Accession: G35070  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-343 <VIR>  
 A:Cross-references: GB:M29008; NID:g192559; PIDN:AAA37414.1; PID:g309165; GB:J05259  
 A:Note: translation of the nucleotide sequence is not complete  
 C:Superfamily: apolipoprotein H; complement factor H repeat homology  
 C:Keywords: duplication  
 F:28-88/Domain: complement factor H repeat homology <FHL>  
 F:92-145/Domain: complement factor H repeat homology <FHR2>  
 F:154-208/Domain: complement factor H repeat homology <FHR3>  
 F:215-269/Domain: complement factor H repeat homology <FHR4>  
 F:273-334/Domain: complement factor H repeat homology <FHR5>  
 Query Match 28.5%; Score 906.5; DB 2; Length 343;  
 Best Local Similarity 32.9%; Pred. No. 9e-55;  
 Matches 187; Conservative 43; Mismatches 99; Indels 239; Gaps 5;  
 QY 1 MLLFSVLISWVSTVGEGTLCDFPKIHGGLYDEEDYNPFSQVPTGEVFFYSCYENFV 60  
 Db 6 LLLLAIVLLTSWFSAKEVSLCDFPKIRHGLYDEKKNPPFSSVLSGKILYSCYENFV 65  
 QY 61 SPKSFWRITRTCTEGSWPTPKCLRMCSFPFVKNGHSSGLIHLEGDTVQIICNTGYSL 120  
 Db 66 SPKSFWRITRTCTEGSWPTPKCLRMCSFPFVKNGHSSGLIHLEGDTVQIICNTGYSL 125  
 QY 121 QNNEKNISCVRGWSTPTPCSTKGECHVPILAEANVDAQPKKESYKGVGLVLFSCRKNLI 180  
 Db 126 QNNQSTITCAEBSWSTPKCISTN----- 149  
 QY 181 RVGSDSVQCYQGWSPNFPCTCKGQVRSCGPPPPQLSNGEVKEIKREYGHNEVVEYDCMNP 240  
 Db 150 -----PTGK-----CGPPPIDNGDITSL----- 168

QY 241 FILINGPKKIQCVGNTTLPCTVEQVTCGYIPELEYGYVQSPVPPYOHGYSVEVNCNE 300  
Db 169 -----SLPVASLSSEYQOCY 186  
QY 301 YAMIGNNMITCINGIWTLPVCMVATHQKCRKCIAGVNIKTLLKLSGKREFNHSRIRYRCS 360  
Db 187 YLLKGNKTIITCRNGKWSPPFCI----- 209  
QY 361 DIFRYRHVSVCINGKWNPEVCTEKREOFCPPPPQIPNAQNMTTNTYNDGKVAVLCKEN 420  
Db 210 -----YPTGK----- 214  
QY 421 YLLPBAKEIVCKDGRWQSLPRCVSTAYCGPPPSINNGDTSFPLSYVPPGSTVYTRCOS 480  
Db 215 -----CGPPPIDNGDITSLSLLEYPLSSVEYQOCN 246  
QY 481 FYKLGQSVTVTCRNKQWSEPPRCIDPCVVSENNKNNIOLKWRNDGKLYAKTGDVAFQ 540  
Db 247 YVVLKGGKTIITCRNGKWSPPFCI----- 306  
QY 541 CKFPHKAMISSPPFAICQEGKFTYPC 568  
Db 307 CKPRYKRAKGLSPERTQINGHINTPTC 334

## RESULT 7

S65551  
factor H - bovine (fragment)  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C:Accession: S65551  
R:Soames, C.J.; Day, A.J.; Sim, R.B.  
Biochem. J. 315, 523-531, 1996  
A:Title: Prediction from sequence comparisons of residues of factor H involved in the in  
A:Reference number: S65551; MUID:96202005; PMID:8615824  
A:Accession: S65551  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA

A:Residues: 1-669 <SOA>  
A:Cross-references: GB:X98697; NID:g1419423; PIDN:CAA67257.1; PID:g1419424  
C:Superfamily: complement factor H; complement factor H repeat homology  
F:55-114/Domain: complement factor H repeat homology <FHR1>  
F:296-349/Domain: complement factor H repeat homology <FHR2>  
F:355-412/Domain: complement factor H repeat homology <FHR3>  
F:416-471/Domain: complement factor H repeat homology <FHR4>  
F:476-530/Domain: complement factor H repeat homology <FHR5>  
F:538-592/Domain: complement factor H repeat homology <FHR6>  
F:599-651/Domain: complement factor H repeat homology <FHR7>

Query Match 27.0%; Score 861; DB 2; Length 669;  
Best Local Similarity 38.1%; Pred. No. 2.6e-51;  
Matches 164; Conservative 54; Mismatches 89; Indels 124; Gaps 5;

QY 23 CDRPKIHGFLYDEEDYNPFSQVPTGEVYVYSCFYNVSPSKSFWTITCTEAGWSTPK 82  
Db 234 CSYPIVIRHGRLY--YSTRGFPPARVNOQFVYSDHHPVPPSQRSDHDLACTAGWSPEEP 291  
QY 83 CLRMCSPPFKNGHSESSGLIHLEGGTVQIICNTGYSLONNEKNISCVVERGWSTPIC-- 140  
Db 292 CLROCIENYLENGHNOHREKYLQGETVVRVICYEGYSLQNDQNTMTCTESGWSPPRCIR 351  
QY 141 -----SPT----- 143  
Db 352 VKTCSKNIRIENGFLSESTFTYPLANKQTEYKCKPGVVTADGKTSGLITCLKNWSAQPV 411  
QY 144 ----- 143  
Db 412 CIKSCDRPVEKARVKSDBGTFWRLNDRLDYECVDGVENRDRGRTGTSVCGDQGSWAKAC 471  
QY 144 -KGECHVPILAEANVDAQPKKESYKVGDLVLFKCRKNLIRVGSQVQCYQFGWSPNFPPTCK 202  
Db 472 YERECSIPEDMPDLNAPPRETYKTVKVGDLVLFKSCSQGRIMYGAUSVQCYHFGWSPKLPPTCK 531

QY 203 -GQVRSQPPQLSNGEYKEIRKEEYGHNEVVEYDCNPNFTINGPKKIQCVDGWTTLPT 261  
Db 532 VKVKSCALPPELPNGKRKEIKHEYAHNEVVEYACNPKFLMKGSHKIQCVDGESTALPV 591  
QY 262 CVEQVKTCGYIPELEYGYVQSPVPPYOHGYSVEVNCREYAMIGNNMITCINGIWTLP 321  
Db 592 CIEBERTIC-ELISLDHGDVPSVPIIHGDSVEFSCREAFMTMIGPRFTICISGEWTPPPQ 650  
QY 322 CVATHQLKRCK 332  
Db 651 CIATDELKRCK 661

## RESULT 8

A46013  
coagulation factor XIII subunit b - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Jun-2000  
C:Accession: A46013  
R:Nonaka, M.; Matsuda, Y.; Shirolahi, T.; Moriwaki, K.; Nonaka, M.; Natsume-Sakai, S  
Genomics 15, 535-542, 1993  
A:Title: Molecular cloning of the b subunit of mouse coagulation factor XIII and assi  
A:Reference number: A46013; MUID:93224141; PMID:8468048  
A:Accession: A46013  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-668 <NON>  
A:Cross-references: GB:D10071; NID:g303651; PIDN:BAA00963.1; PID:g303652  
A:Experimental source: liver

A:Note: sequence extracted from NCBI backbone (NCBI:129405, NCBI:129407)  
C:Superfamily: coagulation factor XIII chain b; complement factor H repeat homology  
F:25-87/Domain: complement factor H repeat homology <FH01>  
F:91-146/Domain: complement factor H repeat homology <FH02>  
F:153-208/Domain: complement factor H repeat homology <FH03>  
F:213-327/Domain: complement factor H repeat homology <FH04>  
F:274-327/Domain: complement factor H repeat homology <FH05>  
F:336-389/Domain: complement factor H repeat homology <FH06>  
F:396-450/Domain: complement factor H repeat homology <FH07>  
F:454-515/Domain: complement factor H repeat homology <FH08>  
F:524-578/Domain: complement factor H repeat homology <FH09>  
F:582-646/Domain: complement factor H repeat homology <FH10>

Query Match 24.6%; Score 783; DB 2; Length 668;  
Best Local Similarity 28.7%; Pred. No. 5.6e-46;  
Matches 187; Conservative 95; Mismatches 266; Indels 104; Gaps 18;

QY 7 VILISWVSTVGGEGTLCDFPKIHG---FLYD-EEDYNPFSQVPTGEVYVYSCYENFVS 61  
Db 9 ILILLSGELYAEKQCDFTVENGRITAQYITFKSFYFPMs---VDKLSFFCLAGYAT 65  
QY 62 PSKSFTRITCTEFGSPPTPKCLRMCSFPFVKNGHSESSGLIHLEGGTVQIICNTGYSLQ 121  
Db 66 ESGKQEQEIRCTAGHSNPRCYKCLKPLDRNGYVSNQKLVKLOERNYSYCGSSGYKTT 125  
QY 122 --NNEKNISCVVERGWSTPICSTFKGECVHPIL- 153  
Db 126 GKDDEEVVHCLSGWSQSPSCREKETCLAPELEHGNYSTQRTFKVKDIAVATCTAGY 185  
QY 154 -----ANVDAQPKKESYKVGDLVLFKCRKNL 179  
Db 186 TTTGKQTGEAECAQNGSLTPQCCKLMCSRLRIENGYPHPVQKQTEEGDLVQFFCHENY 245  
QY 180 IRVGSQVQCYQFGWSPNFPPTCKGQVRSQPPQLSNGEYKEIRKEEYGHNEVVEYDCNP 239  
Db 246 YLSSDILQCYNGWYSPESPICBGRNRNRCPPPPVPLNSKIQP-HSTTYRGERVHIECEL 304  
QY 240 NFIINGPKKIQCVDGWTTLPTCVQVK--TCGYIPELEYGYVQSPVPPYOHGYSVEVNC 297  
Db 305 NFVIQSEILLCELLCNGKWTPEPKCIEBKVACQPPSVENGVAHPHSEIYVYSGDKVYTRC 364  
QY 298 RNEYAMIGNNMITCINGIWTLPVCMVATHQKCRCK---IA-CVNITKTLKLSGKREFNHN 352  
Db 365 GGSYLRSGSSTITCNRGRWTLPPCEV--ENIENCKPPPPDIANGVVVDGLL----ASYTTG 418





F; 28-81/Domain: complement factor H repeat homology <FH1>  
F; 86-146/Domain: complement factor H repeat homology <FH2>  
F; 150-203/Domain: complement factor H repeat homology <FH3>  
F; 209-266/Domain: complement factor H repeat homology <FH4>  
F; 270-325/Domain: complement factor H repeat homology <FH5>  
F; 332-386/Domain: complement factor H repeat homology <FH6>  
F; 390-451/Domain: complement factor H repeat homology <FH7>

Query Match 20.3%; Score 646; DB 2; Length 452;  
Best Local Similarity 24.9%; Pred. No. 8.5e-37;  
Matches 157; Conservative 66; Mismatches 160; Indels 248; Gaps 1

QY 1 MLLFSVILISWSTVGGEG-----20  
DB 6 MLLLSNILLTAWLSTAKGEKTCSPPIYLINGIYTTHRIHSDDIRVECNWGFVPVTGS 65  
QY 21 -----TL--CDPPKIHGHFLYDEEDYNPFSSQVPTGTFYFYSCYENFVSP 62  
DB 66 TVSKCTPTGWIPVRCTLKPCFPQFKYGLRYEESLRPNRPVSIGNKYSYRCDNGFSP 124  
QY 63 SKSFWRITICTEGNSPTPKCLRMCSFFPVKNHGHSESSLIHLEODTVQIIICNTGYSLON 124  
DB 126 SGYSWDYLRCTAGQEPEVCYRKCVFHYHVENGDSAYWEKIYVGQSLLKVQCYNGYSLON 181  
QY 123 NEKNISCVRKGWSTPICSTFKGECVHPILEANVDQAQPKESYKVGDDVLKFCRKNLTRV 181  
DB 186 GDIMICTENGWSPPKC-----KTCAS-----DIHDNGFLSE--SSYIALNRETS YRCKQGYY 244  
QY 183 GDSVOCYOFGWSPNPPTCKGOVRSCGPPQLSNGEVKEIRKEEYGHNEVVEYDCNPNFI 244  
DB 207 -----KTCAS-----DIHDNGFLSE--SSYIALNRETS YRCKQGYY 244  
QY 243 INCPK----KIQCVDGEWITLTPTCEQVKTCGYIPELEYGVQPSVPPQHGVSVVEVNCRN 299  
DB 244 TMTGETSGITCLONGWSQPSC---IKSC-----DMPVFENSIT-----280  
QY 300 EYAMIGNNMITCINGIWTELPNCVATHQLKRCKIAGVNIKTLLKSGEFWNHSRIYRC 353  
DB 281 -----KTRTWFKL-----NDKLDEC 297  
QY 360 SDIF--RYRHVSICNGKANPEVDCTEKREQFCPPPIQIPNAONMTTVVYODGEKVAVLC 411  
DB 298 LVGFENEYKHT-----LVGFENEYKHT-----308  
QY 418 KENYLLPKEAKEIVCKDRWQSLPRCVESTAYCGPPPSINNGDTTSFPLSYPPGSTVTYR 477  
DB 309 -----KGSITCYTGMSDTPSCYDSTRYCGPPPIDNGDITSLSLPEYELUSSVDYO 360  
QY 478 CQSFYKLGQSVTVTCRNKOWSEPRLCPDQCVVSVBNNMKNIQLKWNRDGLKYAKTGDAV 533  
DB 361 CQKYLYLLGNKNTITCRNGEKWSEPPCLHACVIPENIMEAHNILKWRHTEKIYAHSGEDI 420  
QY 538 EFQCFPHKAMISSPPFRACOEGKFEPYC 568  
DB 421 EPECRGYOKARGSLPFTKCTINGTINYPTC 451

RESULT 12  
S46199  
probable complement regulatory plasma protein SBI - barred sand bass  
C:Species: Paralabrax nebulifer  
C:Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 22-Mar-2002  
C:Accession: S46199; S77894  
C:Dahmen, A.; Kaidoh, T.; Zipfel, P.F.; Gigli, I.  
Biochem. J. 301, 391-397, 1994  
A:Title: Cloning and characterization of a cDNA representing a putative complement factor H from the sea bream, *Lepomis microlophus*.  
A:Reference number: S46199; MUID:94318039; PMID:8042982  
A:Accession: S46199  
A:Molecule type: mRNA  
A:Residues: 1-1053 <DAH1>  
A:Cross-references: EMBL:L21703; NID:g639894; PIDN:AA92556.1; PID:g639895  
A:Experimental source: liver  
A:Accession: S77894





Db 1362 CELP-VHAGHCKTPEOF-PRAS-PTIPINDFEFPVGTSLNYECRPGYFGAMFISCLLENL 1418  
Qy 60 VSPSKSFWRIT--TCTBEGMSPTPKCLRMCSFFPVKNHSESSGLIHL-----GDTVQI 112  
Db 1419 V-----WSSVEDNCRKSCGPPPE-----PF-----NGVHINTDQFGSTVNY 1457  
Qy 113 ICNTGYSL-----QNN-----123  
Db 1458 SCNEGRFLIGSPSTCLVSGNNVTWKKAPICEIISCEPPPTISNGDFYNNRASFNHGT 1517  
Qy 124 -----EKNISCVERG-----WSTPTICSTFKGECHVPPILEANVDA 158  
Db 1518 VTYOCHTGDGEQLFELVGERISYCTSKDDQGVWSSPPRCISTNKCTAPEVNAIRV 1577  
Qy 159 QPKKESYKVGDLVKFKCRKKLIRVSDSVQCYQFG--WSPNFPCTKGQVRS CGPPPOLNSG 217  
Db 1578 PGNRSFSLTEIVFRFCQFVWVGSHTVOCQTNRWRGPKLPHCS---RVCQPPPEILHG 1634  
Qy 218 EVKEIRKEEYGHNEVEYDCNPNFIINGPKKIQCV-DGEWT-TLPTCQEOKTC-GYIPE 274  
Db 1635 EHTPSHODNFSQGEVYFCEPGYDLRGAASLHCTPOGDWSPAPRCT--VKSCDDFLGQ 1692  
Qy 275 LEYGYVQPSVPPYQHGVSFVNCRNEYAMIGNNMITCINGIWTLELPMCVATHOLKCKIA 334  
Db 1693 LPHGRVLFPL-NLQLGAKVSFVCDGEGFRLKGRS-----ASH-----CVLA 1731  
Qy 335 GVNIKTLLKLSGREFNHSRIYRCSDIFRYHSVCINGKNPVEVDCTEKREOCPCPPPO 394  
Db 1732 G--MKAL-----WNSVVPCE--QIFCPNPPA 1754  
Qy 395 IPNAQNMTT---TVNYODGKVAVLCKEN-----YLLPEAKEIVCK-----DGRWQS-L 439  
Db 1755 ILNGRHGTPTFGDIPY--GKEISYACDTHPDGRGMTFNLIGESSTRICTSDPQNGVWSSPA 1812  
Qy 440 PRCVESPT-AYCGPPPSINNGDTSFPLSVYPPGCVTVTYRCOSFYKLOGSVTVTCRNKQ-W 497  
Db 1813 PRCELSVPAACPHPPKIQNGHDIGGHVSLYLPGMTISYICDPGYLLVKGFIPTDQGIW 1872  
Qy 498 SE-----PPRCLDPCVVSEENMNKNIQLKWRNDGKLYAKTGDAVEFOCK 542  
Db 1873 SOLDHCKEVCNCEPLF-----MNGISKELEMK---KVY-HYGDYVTLKCE 1914

Search completed: July 17, 2003, 13:54:23  
Job time : 48 secs

**THIS PAGE BLANK (USPTO)**